Cryptographic Hardness based on the Decoding of Reed-Solomon Codes

Aggelos Kiayias* Moti Yung †

Abstract

We investigate the decoding problem of Reed-Solomon (RS) Codes, also known as the Polynomial Reconstruction Problem (PR), from a cryptographic hardness perspective. Namely, we deal with PR instances with parameter choices for which decoding is not known to be feasibly solvable and where part of the solution polynomial is the hidden input. We put forth a natural decisional intractability assumption that relates to this decoding problem: distinguishing between a single randomly chosen error-location and a single randomly chosen non-error location for a given corrupted RS codeword with random noise. We prove that under this assumption, PR-instances are entirely pseudorandom, i.e., they are indistinguishable from random vectors over the underlying finite field. Moreover, under the same assumption we show that it is hard to extract any partial information related to the hidden input encoded by the corrupted PR-instance, i.e., PR-instances hide their message polynomial solution in the semantic security sense.

The above results lay a framework for the exploitation of PR as an intractability assumption for provable security of cryptographic primitives. Based on this framework, we present provably secure cryptographic constructions for (i) a pseudorandom number generator, (ii) a semantically secure version of the Oblivious Polynomial Evaluation Protocol, and (iii) a stateful cipher with a set of interesting properties that include: semantic security, forward secrecy, error-correcting decryption and an array of random self-reducibility properties with respect to the plaintext choice, key choice and partial domain choice.

1 Introduction

Finding new problems based on which we can design cryptographic primitives is an important research area. Given a presumably hard problem, it is usually non-trivial to exploit it directly in cryptography. In fact, many times in order to serve as the base for secure cryptographic primitives, we need to find related hard decision problems (predicates). This is the fundamental methodology initiated by Goldwasser and Micali in [GM84] where they started the quest for formal notions and proofs of security in cryptography. The decision problem's hardness, typically seems related to (or at times proved in some sense related or, even better, reducible from) the hardness of the original problem. Hard predicate assumptions allow formal security proofs (in the form of reductions) for advanced cryptographic primitives such as pseudorandomness and semantically secure encryption. The first example of a decisional assumption is Quadratic-Residuosity, which is related to (but not known to be reducible from) Factoring and

*University of Connecticut, Storrs, CT, USA, aggelos@cse.uconn.edu
†RSA Laboratories and Columbia University, USA, moti@cs.columbia.edu

was employed in designing the first semantically secure encryption scheme [GM84]. Another such assumption is the Decisional Diffie-Hellman which implies the security of ElGamal encryption and other advanced cryptographic primitives (e.g., [NR98]), and is related to (but not known to be reducible from) the Diffie-Hellman problem.

In this work, our goal is to investigate the possibility of cryptographic primitives whose security is based on the problem of *Polynomial Reconstruction* (PR). Recall that the problem of Polynomial Reconstruction is defined as follows: Given n points over a finite field $\mathbb{F}$, such that at least t of them belong to the graph of a polynomial p of degree less than k, recover such a polynomial (where $n > t > k$).

We note that Polynomial Reconstruction is equivalent to the decoding problem of Reed-Solomon codes [RS60] and naturally has received much attention from a "positive" (coding theoretic) perspective: Starting from the classical algorithm of Berlekamp and Welch ([BW86]) which solves Polynomial Reconstruction provided that $t \geq \frac{n+k}{2}$ (which matches the error correcting bound for Reed-Solomon Codes), to the recent work of Guruswami and Sudan [GS98] which solves it when $t \geq \sqrt{kn}$ (where many solutions are possible in the worst case). The current state of knowledge suggests that for values of t below $\sqrt{kn}$ the problem may be hard (even under the light of recent extensions of list or average case decoding for related families of codes in [BKY03, CS03, PV05]).

Regarding our goal, Polynomial Reconstruction as is does not appear to be amenable to direct cryptographic exploitation: even if presumed hard, it is not at all clear how to build advanced cryptographic primitives whose security can be reduced to it. Indeed, when Naor and Pinkas [NP99] first employed the problem cryptographically in a context of protocol design, they introduced a related pseudorandomness assumption. The relation of this assumption to PR is another motivation for further investigation.

In this work, we identify a decisional problem that is naturally related to PR. The decisional problem is defined as a distinguishability challenge between two ensembles (families of distributions): the first ensemble, contains pairs of the form $(i, \mathbf{y})$ where $\mathbf{y}$ is a random PR-instance and i is a random error-location of that instance, whereas the second contains similar pairs where i is a random non-error-location of the given instance. A distinguisher solves the decisional problem if it can tell the two ensembles apart with some substantial advantage. The indistinguishability of the two distributions is what we call "Decisional-PR-Assumption" (DPR).

The DPR as formulated in the present work is a natural assumption and appears to be intimately related to the decoding problem: the task of any decoder is to distinguish between error and non-error locations. With thios assumption as a starting point, we then proceed to employ the DPR in the cryptographic setting.

Pseudorandomnessof PR instances: We first prove that the DPR implies that PR-instances are pseudorandom, i.e., they are computationally indistinguishable from random vectors. This result is the fundamental backbone of the present investigation as pseudorandomness of PR-instances is quite amenable to cryptographic exploitation as a reduction basis: indeed, the security of all other constructions in this work will be reducible from pseudorandomness that in turn is reducible from DPR.

Hardness of partial information extraction: We then show that an adversary with access to a PR-instance who wishes to predict the value of some computable function on a portion of the polynomial solution has only negligible advantage. This holds true even if the portion of

the polynomial solution curve follows an adversarially chosen probability distribution. This suggests that, under DPR a PR-instance semantically hides portions of its polynomial solution.

The above results lay the framework for the applied cryptographic exploitation of the PR problem in provable security. The advantages of basing cryptographic constructions on a PR related assumption, besides the fundamental insight into the PR problem, are as follows: (i) the basic operation of PR-based cryptographic primitives is polynomial interpolation which can be implemented quite efficiently (compared to e.g., modular exponentiation or other cryptographic operations). (ii) PR seems to be a hard problem, and in the worst-case a variant of the problem has been shown to be NP-hard [GSR95]; this may suggest that it could be difficult to solve PR even with the advent of quantum computation, cf. [BBBV97].

We present the following cryptographic applications.

Pseudorandomness Extender. We design a bitstring mapping that given a seed it extends it to longer bitstring that appears to be random to any polynomial time bounded attacker under the DPR assumption. Based on such pseudorandom extension of an initial random seed, it is straightforward to derive a pseudorandom number generator, cf. Chapter 3 of [Gol01].

Semantic Security for Oblivious Polynomial Evaluation. In [NP99, NP06], the primitive of oblivious polynomial evalutation (OPE) was introduced: an OPE protocol allows a party A to obtain the point evaluation on a polynomial held by a party B, so that A reveals no information about the point she evaluates and B reveals no information about his polynomial beyond the point that A learns. It is possible to implement OPE based on an t-out-of-n oblivious transfer by having party A send her point encapsulated into a noisy PR-instance ([NP99, NP06], see also [KY04]). In this work we show that the PR based protocol for OPE satisfies semantic security for player A assuming the DPR assumption.

Stateful Cipher. We design a stateful block cipher that enables secure communication between two parties and satisfies a number of interesting properties: (i) *semantic security*: an adaptive chosen plaintext attacker does not get any advantage in guessing any non-trivial property of a given challenge ciphertext, (ii) *forward secrecy*: if a total security breach occurs at a certain time (e.g. the key is revealed), this affects the security only of future messages while the previously sent messages are semantically secure in the view of the perpetrator. (iii) *Random self-reducibility properties*: A typical security concern is whether an attacker may take advantage of the structure of a subset of the key-space or of a subset of the plaintext-space. We show that an attack on a large enough subset is equivalent to an attack on the average case, for both plaintext space and key space, showing there are no weak plaintext subset and no weak key subsets. In a similar fashion, we can also define random self-reducibility with respect to a partial domain function: in this case we need to transform an adversary that recovers a certain portion of the plaintext to an adversary that attacks a different portion of the plaintext. Intuitively this means that no particular portion of the plaintext is more advantageous to attack. We show that the PR-cipher satisfies this random self-reducibility property as well. We remark that such properties are satisfied unconditionally based on the problem's inherent structure (as opposed to e.g., the semantic security of the cipher). The final property that the cipher possesses due to its underlying nature is (iv) *error-correcting decryption*: the decryption operation incorporates error-correction capabilities in a direct manner.

Remark: The present work is a full revised version of [KY02]. It contains a novel formulation of the decisional PR assumption (that is much more natural compared to the original formulation), a number of corrections and a complete restructuring of the exposition of the security

proofs.

Notation. All computations are performed in a (large) finite field $\mathbb{F}$ of prime order. Tuples in $\mathbb{F}^n$ are denoted by $\mathbf{x}$ and $(\mathbf{x})_i$ denotes the i-th coordinate of $\mathbf{x}$. For commonly used tuples such as $\mathbf{z}, \mathbf{y}$ we will also use the notation z_i, y_i for $(\mathbf{z})_i, (\mathbf{y})_i$, respectively. PPT stands for "probabilistic polynomial-time." All algorithms mentioned in the paper are polynomial-time Turing Machines, and denoted by $\mathcal{A}, \mathcal{B}$ etc. For any PPT $\mathcal{A}$ that uses randomness $r \in \mathcal{R}$ and has input distributed according to some distribution $\mathcal{D}$, if y is in the range of $\mathcal{A}$ we will denote by $\mathsf{Pr}[\mathcal{A}(r, x) = y : r \leftarrow \mathcal{R}, x \leftarrow \mathcal{D}]$ the probability that $\mathcal{A}$ returns y when its input is distributed according to $\mathcal{D}$; occasionally we may drop r from the above notation. A function $\alpha(\lambda) : \mathbb{N} \to \mathbb{R}$ is negligible if for all c it holds that $\alpha < \lambda^{-c}$ for sufficiently large λ. If the probability of an event is greater equal to $1 - \epsilon$ where $\epsilon(\lambda)$ is a negligible function, then we write that the event happens "with overwhelming probability."

2 PR as a Cryptographically Hard Problem

Definition 2.1 Polynomial Reconstruction (PR). *Given n, k, t and $\mathbf{z}, \mathbf{y} \in \mathbb{F}^n$ with $z_i \neq z_j$ for $i \neq j$, output all $\langle p(x), I\rangle$ such that $p \in \mathbb{F}[x]$, $\text{degree}(p) < k$, $I \subseteq \{1, \ldots, n\}$, $|I| \geq t$ and $\forall i \in I(p(z_i) = y_i)$.*

PR as a coding theoretic problem asks for all messages that agree with at least t positions of the received Reed-Solomon codeword. For a general treatment on the subject the interested reader is referred to [Ber68] or [MS77]. Note that $k < n$ since k/n is the message rate of the code, and that we further require that at least one solution $\langle p(x), I\rangle$ exists.

When $t \geq \frac{n+k}{2}$ then PR$[\mathbf{z}, k, t]$ has only one solution and it can be found with the algorithm of Berlekamp and Welch [BW86] ($\frac{n+k}{2}$ is the error-correction bound of the Reed-Solomon codes). When t is beyond the error-correction bound then having more than one solution is possible. Sudan proposed an algorithm that solves the PR beyond the error-correction bound when $t \geq \sqrt{2kn}$ in [Sud97] and later in [GS98], Guruswami and Sudan presented an algorithm that solves the PR for $t > \sqrt{kn}$. In [GSR95] it was proven that when $t > \sqrt{kn}$ the number of solutions is bounded by a polynomial. In [GS98] it is pointed out that the possibility of an algorithm that solves instances for smaller values of t might be limited. Consequently the current state of knowledge implies that PR$[\mathbf{z}, k, t]$ is hard for the choice of parameters $t < \sqrt{kn}$.

Input Convention for PR algorithms. A PR-instance will be denoted by $\mathbf{y}$ and defined for the parameters $n, k, t, \mathbf{z}$. When we say that an algorithm $\mathcal{A}$ operates with input a PR instance we will simply write $\mathcal{A}(\mathbf{y})$ instead of $\mathcal{A}(n, k, t, \mathbf{z}, \mathbf{y})$.

2.1 Structure of the Instance Space

An instance of PR is a vector $\mathbf{y} = \langle y_1, \ldots, y_n\rangle \in \mathbb{F}^n$ and is specified by the parameters n, k, t and the support elements $\mathbf{z} = \langle z_1, \ldots, z_n\rangle \in \mathbb{F}^n$. Note that the support elements are all distinct but are given in vector form to define their correspondence to the elements of the vector $\mathbf{y}$.

Let $\mathbf{e} \in \mathbb{F}^n$ be vector of Hamming weight at most $n - t$. The general structure of a PR-instance with parameters n, k, t and support $\mathbf{z}$ that we consider is $\mathbf{y} = \mathbf{e} + \mathbf{p}$ where $p \in \mathbb{F}[x]$, and $\mathbf{p} = \langle p(z_1), \ldots, p(z_n)\rangle$. The set of all PR-instances with parameters n, k, t over the support $\mathbf{z}$ will be denoted by $\mathcal{I}^{\mathbf{z}}_{n,k,t}$.

From this point we will denote by $\mathbf{y}_1, \mathbf{y}_2, \ldots$ arbitrary elements of $\mathbb{F}^n$, by $\mathbf{e}_1, \mathbf{e}_2, \ldots$ arbitrary error vectors, and by $\mathbf{p}_1, \mathbf{p}_2, \ldots$ vectors of the form $\langle p(z_1), \ldots, p(z_n)\rangle$ such that $p \in \mathbb{F}[k]$ and the degree of p is less than k. The parameters $n, k, t, \mathbf{z}$ will be clear from the context.

We know that in case $t \geq \frac{n+k}{2}$, independently of the choice of $\mathbf{e}$, the PR-instance $\mathbf{y} = \mathbf{e} + \mathbf{p}$ has the single unique solution $p \in \mathbb{F}[x]$. Nevertheless, for smaller values of t it may be the case that a PR-instance $\mathbf{y}$ may have more than one solutions. This is the setting when we will say that the error-vector is ambiguous. Consider the following definition:

Definition 2.2 *A vector $\mathbf{e} = \langle e_1, \ldots, e_n\rangle$ is an (n,t)-*error-vector for PR-instances with parameters n, k, t *if its Hamming weight is $n - t$. An (n,t)-error-vector $\mathbf{e}$ is called k-*ambiguous *if there is a polynomial $p \neq 0$ of degree less than k and a set of indices I with $|I| = t$ so that $p(z_i) = e_i$ for all $i \in I$.*

Lemma 2.3 *Suppose $\mathbf{y} \in \mathcal{I}^{\mathbf{z}}_{n,k,t}$; the following two statements are equivalent:*

- *There exist $\mathbf{p}, \mathbf{e}$ such that $\mathbf{y} = \mathbf{p} + \mathbf{e}$ and $\mathbf{e}$ is a k-ambiguous (n,t)-error-vector.*
- *There exist $\mathbf{p}_1, \mathbf{e}_1, \mathbf{p}_2, \mathbf{e}_2$ such that $\mathbf{y} = \mathbf{p}_1 + \mathbf{e}_1 = \mathbf{p}_2 + \mathbf{e}_2$ and $\mathbf{e}_1 \neq \mathbf{e}_2$.*

Proof. Suppose that $\mathbf{y} = \mathbf{p} + \mathbf{e}$ and $\mathbf{e}$ is k-ambiguous. This means that there exists some non-zero polynomial $p' \in \mathbb{F}[x]$ of degree less than k such that $(\mathbf{e})_i = p'(z_i)$ for all $i \in I'$ where I' is a set of indices from $\{1, \ldots, n\}$ of size t. Consider the vector $\mathbf{e}' = \mathbf{e} - \mathbf{p}'$. It is easy to see that it is an (n,t)-error-vector. Moreover it holds that $\mathbf{y} = (\mathbf{p} + \mathbf{p}') + \mathbf{e}'$. Note that we have that $\mathbf{e} \neq \mathbf{e}'$ since $p' \neq 0$. This completes the first step of the proof.

Suppose now that $\mathbf{y} = \mathbf{p}_1 + \mathbf{e_1} = \mathbf{p}_2 + \mathbf{e}_2$ with $\mathbf{e}_1 \neq \mathbf{e}_2$. It follows that $\mathbf{e}_1 = \mathbf{e}_2 + (\mathbf{p}_2 - \mathbf{p}_1)$; given that $\mathbf{e}_2$ is an (n,t)-error-vector we have that there is a set of indices I_2 of size t such that $(\mathbf{e}_2)_i = 0$ for all $i \in I_2$. Based on this we have that $(\mathbf{e}_1)_i = (p_2(z_i) - p_1(z_i))$ for $i \in I_2$. Note that the polynomial $p_2 - p_1$ must be non-zero (because if it is zero then it holds that $\mathbf{e}_1 = \mathbf{e}_2$, a contradiction). Moreover it is clear that the degree of $p_2 - p_1$ is less than k. This shows that $\mathbf{e}_1$ is k-ambiguous and completes the second step of the proof. ■

Next we show some basic results about the number of error-vectors in our formulation:

Lemma 2.4 *(1) The total number of (n,t)-error-vectors is equal to $\binom{n}{t}(|\mathbb{F}| - 1)^{n-t}$.*
(2) The number of (n,t)-error-vectors that are ambiguous is less than $\binom{n}{t}^2(|\mathbb{F}| - 1)^{n-2t+k}$.

Proof. (1) The formula is straightforward, since the vector contains $n - t$ non-zero points that can be distributed in $\binom{n}{n-t}$ ways.

(2) An ambiguous vector by definition contains an embedded non-zero polynomial p that matches it in (at least) t locations. Note that such locations may span both the zero as well as non-zero areas of the error-location. Let $m \in \{0, \ldots, k-1\}$ be a value that specifies in how many of the zero locations such polynomial p will match the error vector (note that it cannot be that $m = k$ since then it holds that $p = 0$). The number of ways we can do select an ambiguous error-vector will be at most:

$$\sum_{m=\max\{2t-n,0\}}^{k-1} \binom{n}{t}\binom{t}{m}\binom{n-t}{t-m}(|\mathbb{F}| - 1)^{(n-t)-(t-m)+(k-m)}$$

This is because, $\binom{n}{t}$ is the number of ways to choose t non-zero locations among n points, $\binom{t}{m}$ is the number of ways we may choose the zero locations over which the polynomial p will cross, $\binom{n-t}{t-m}$ is the number of ways we may choose the remaining non-zero locations over which the polynomial p will pass. Finally, we have $n-t-(t-m)+(k-m)$ non-zero points: this is because there are $t-m$ of the non-zero error-points that are not freely selected but controlled by the polynomial which can contribute only $k-m$ degrees of freedom.

The bounds of the summation are justified as follows: m should be at least $2t-n$ unless this quantity drops below zero where in this case we simply start m from zero. Note that $2t-n<k$ since the case $t \geq \frac{n+k}{2}$ is excluded as there cannot be ambiguous error vectors for such parameter choices. Simplifying the above summation we obtain that it is at most $\binom{n}{t}^2(|\mathbb{F}|-1)^{n-2t+k}$. ■

Corollary 2.5 *Suppose* $\log(|\mathbb{F}|-1) \geq (\log\binom{n}{t}+s)(t-k)^{-1}$*; then, (i) the probability of a uniform random variable over all* (n,t)*-error-vectors to result in an ambiguous* (n,t)*-error-vector is less than* 2^{-s}
(ii) $(1-2^{-s}) \leq \frac{|\mathcal{I}^{\mathbf{z}}_{n,k,t}|}{|\mathbb{F}|^k\binom{n}{t}(|\mathbb{F}|-1)^{n-t}} \leq 1$.

Proof. (i) Follows directly from lemma 2.4 since $\binom{n}{t}^2(|\mathbb{F}|-1)^{n-2t+k}/\binom{n}{t}(|\mathbb{F}|-1)^{n-t}$ equals $\binom{n}{t}(|\mathbb{F}|-1)^{-t+k}$ which is less or equal to 2^{-s} based on the given condition.
(ii) Let $\mathcal{I}^{\mathbf{z}}_{n,k,t}(I)$ be the subset of $\mathcal{I}^{\mathbf{z}}_{n,k,t}$ that contains vectors $\mathbf{y}$ so that we can decompose $\mathbf{y}=\mathbf{p}+\mathbf{e}$ and $\mathbf{e}$ is a (n,t)-error-vector that has zero's in the locations I. It follows that all instances will be included in the set $\cup_I \mathcal{I}^{\mathbf{z}}_{n,k,t}(I)$, where I is selected at from the set of all subsets of size t of $\{1,\ldots,n\}$. The union bound immediately yields the rightmost inequality of the statement (ii) above. For the other leftmost inequality, observe that the number of non-ambiguous (n,t)-error-vectors is at least $\binom{n}{t}(|\mathbb{F}|-1)^{n-t}-\binom{n}{t}^2(|\mathbb{F}|-1)^{n-2t+k}$ using lemma 2.4. Based on this we deduce that $|\mathcal{I}^{\mathbf{z}}_{n,k,t}| \geq |\mathbb{F}|^k(\binom{n}{t}(|\mathbb{F}|-1)^{n-t}-\binom{n}{t}^2(|\mathbb{F}|-1)^{n-2t+k})$. From this we obtain that $|\mathcal{I}^{\mathbf{z}}_{n,k,t}|/(|\mathbb{F}|^k\binom{n}{t}(|\mathbb{F}|-1)^{n-t}) \geq (1-\binom{n}{t}/(|\mathbb{F}|-1)^{t-k})$ from which the statement of the corollary follows. ■

Given that we will be interested in sampling "hard" instances of $\mathrm{PR}[\mathbf{z},k,t]$ we will specify next an efficient sampler that will be used for this purpose:

Definition 2.6 (Sampling PR-instances) *Consider the following sampler* S *that produces an instance of* $\mathcal{I}^{\mathbf{z}}_{n,k,t}$*:* S *on input* $(\mathbf{z},k,t)$ *samples a random subset* $I \subseteq \{1,\ldots,n\}$*, a polynomial* $p \in \mathbb{F}[x]$ *with* $\mathrm{degree}(p) < k$*; it then sets* $y_i = p(z_i)$ *for all* $i \in I$*, whereas for all* $i \notin I$ *it samples* y_i *at random from the set* $\mathbb{F}-\{p(z_i)\}$*.* S *terminates by returning the vector* $\mathbf{y}=\langle y_1,\ldots,y_n\rangle$*. When* $\mathbf{y}$ *is the output of* S *we will denote by* $I_{\mathbf{y}}$ *the index set* I*.*

An equivalent description of the above sampler S can also be described as selecting a random polynomial $p \in \mathbb{F}[x]$ with $\mathrm{degree}(p) < k$, as well as a random (n,t)-error-vector $\mathbf{e}$ and returns $\mathbf{p}+\mathbf{e}$. The distribution induced by S over the space of PR-instances $\mathcal{I}^{\mathbf{z}}_{n,k,t}$ will be denoted by $\mathsf{S}^{\mathbf{z}}_{n,k,t}$.

Theorem 2.7 *Suppose* $\log(|\mathbb{F}|-1) \geq (\log\binom{n}{t}+s)(t-k)^{-1}$*; The statistical distance between the distribution* $\mathsf{S}^{\mathbf{z}}_{n,k,t}$ *and the uniform over* $\mathcal{I}^{\mathbf{z}}_{n,k,t}$ *is less than* 2^{-s+2}*.*

Proof. Given a $\mathbf{y} \in \mathcal{I}^{\mathbf{z}}_{n,k,t}$ define $n_{\mathbf{y}}$ to be the number of different pairs $(\mathbf{p}, \mathbf{e})$ that satisfy $\mathbf{y} = \mathbf{p} + \mathbf{e}$. If there is only a single pair that fits $\mathbf{y}$, then we say that $\mathbf{y}$ has a unique solution; denote the set of all $\mathbf{y}$ that have a unique solution by U; denote by $\overline{U}$ all the remaining tuples.

First consider the following summation $A = \sum_{\mathbf{y}\in\overline{U}} |\mathcal{I}^{\mathbf{z}}_{n,k,t}|^{-1} = |\overline{U}|/|\mathcal{I}^{\mathbf{z}}_{n,k,t}|$. Note that $|\overline{U}| < |\mathbb{F}|^k \binom{n}{t}^2 (|\mathbb{F}|-1)^{n-2t+k}$, using lemma 2.4(ii). Next using corollary 2.5(ii) we have that $|\mathcal{I}^{\mathbf{z}}_{n,k,t}| \geq (1-2^{-s})|\mathbb{F}|^k \binom{n}{t} (|\mathbb{F}|-1)^{n-t}$. Based on this we obtain that $A < \binom{n}{t}/(|\mathbb{F}|-1)^{t-k}(1-2^{-s}) \leq 2^{-s}/(1-2^{-s}) = 1/(2^s-1) \leq 2^{-s+1}$.

Next we consider the summation $B = \sum_{\mathbf{y}\in\overline{U}} n_{\mathbf{y}}/C$ where $C = |\mathbb{F}|^k \binom{n}{t} (|\mathbb{F}|-1)^{n-t}$. Observe that $\sum_{\mathbf{y}\in\overline{U}} n_{\mathbf{y}} = C - \sum_{\mathbf{y}\in U} n_{\mathbf{y}} = C - |U|$. Based on this we have that $B = 1 - |U|/C$. Now $|U| \geq |\mathcal{I}^{\mathbf{z}}_{n,k,t}| - |\mathbb{F}|^k \binom{n}{t}^2 (|\mathbb{F}|-1)^{n-2t+k}$. As a result, $|U|/C \geq |I|/C - 2^{-s} \geq 1 - 2^{-s+1}$ using corollary 2.5(ii). It follows that $B \leq 2^{-s+1}$.

The statistical distance of the two distributions equals $\frac{1}{2}\sum_{\mathbf{y}} |\frac{n_{\mathbf{y}}}{C} - |\mathcal{I}^{\mathbf{z}}_{n,k,t}|^{-1}| \leq \frac{1}{2}(\sum_{\mathbf{y}\in U} |C^{-1} - |\mathcal{I}^{\mathbf{z}}_{n,k,t}|^{-1}| + A + B) < \frac{1}{2}(1 - |\mathcal{I}^{\mathbf{z}}_{n,k,t}||/C + 2^{-s+1} + 2^{-s+1}) < 2^{-s+2}$. ■

2.2 Parameter Selection

In our exposition we will use λ as the security parameter. The parameters $n, k, t, \log|\mathbb{F}|$ will all be functions of λ, and will be assumed to satisfy the inequality $k < t < n < |\mathbb{F}|$ as well as $t < \sqrt{nk}$. The straightforward brute-force algorithm for solving $\mathrm{PR}[\mathbf{z}, k, t]$ requires checking all possibilities and as a result, it has complexity proportional to $\min(\binom{n}{k}, \binom{n}{t})$. The parameter selection $[n, k, t]$ would be suitable for the $\mathrm{PR}[\mathbf{z}, k, t]$, if k, t are chosen so that $t < \sqrt{kn}$ and $\min(\binom{n}{k}, \binom{n}{t})$ is exponential in λ. Note that the size of $\mathbb{F}$ can be made arbitrarily large (as long as it is larger than n); regarding the structure of $\mathbb{F}$, we choose $\mathbb{F}$ to be a prime field.

2.3 The Intractability Assumption

A decision problem that relates naturally to the hardness of solving an instance $\mathbf{y}$ of $\mathrm{PR}[\mathbf{z}, k, t]$ is the following: given $\mathbf{y}$ and an index $i \in \{1, \ldots, n\}$ decide whether $i \in I_{\mathbf{y}}$. We will postulate that such decision is computationally hard to make. A natural way to formalize this as an intractability assumption is to define first the following two samplable probability distributions:

Definition 2.8 *Given parameters n, k, t, the sampler $\mathsf{S}^{\mathsf{bad}}$ first selects an instance $\mathbf{y}$ following the sampler S of definition 2.6, then it selects i at random from the set $\{1, \ldots, n\} - I_{\mathbf{y}}$ and then outputs $\langle i, \mathbf{y}\rangle$. $\mathsf{S}^{\mathsf{good}}$ is defined similarly but i is selected at random from the set $I_{\mathbf{y}}$ instead.*

The above two samplers will be used to define the challenge for our decisional PR assumption. In particular we have that:

Definition 2.9 Decisional-PR-Assumption. $\mathrm{DPR}[\mathbf{z}, k, t]$. *For any* PPT $\mathcal{A}$ *we define:*

$$\mathsf{Adv}^{\mathsf{dpr},\mathcal{A}}_{\mathbf{z},k,t}(\lambda) = |\mathsf{Pr}[\mathcal{A}(\mathsf{S}^{\mathsf{good}}(\mathbf{z}, k, t)) = 1] - \mathsf{Pr}[\mathcal{A}(\mathsf{S}^{\mathsf{bad}}(\mathbf{z}, k, t)) = 1]|$$

It holds that $\mathsf{Adv}^{\mathsf{dpr}}_{\mathbf{z},k,t}(\lambda) = \max_{\mathcal{A}} \mathsf{Adv}^{\mathsf{dpr},\mathcal{A}}_{\mathbf{z},k,t}(\lambda) = \mathsf{negl}(\lambda)$.

The intuition behind the assumption is that if the adversary is presented with a challenge location then it is incapable of telling the difference between a good (i.e., uncorrupted) location and a bad (i.e., error) location. The reader should note that that the random variable i defined as the first component in an $\langle i, \mathbf{y} \rangle$ pair drawn from $\mathsf{S}^{\mathsf{good}}$ (or $\mathsf{S}^{\mathsf{bad}}$) is in fact uniformly distributed over $\{1, \ldots, n\}$ (i.e., an adversary concentrating on i itself without considering $\mathbf{y}$obtains no information about the source of the challenge $\langle i, \mathbf{y} \rangle$).

Note that in a previous version of this paper, [KY02], we formulated the DPR assumption differently employing what was called there a gap-predicate-pair. Such a pair of PPT predicates exhibited a non-negligible spike for some specific location that had to be necessarily one of the good locations; the DPR then suggested that gap-predicate-pairs do not exist. Our present formulation is more intuitive in comparison and more general (in particular it can be shown easily that the DPR as formulated above would be violated if any a gap-predicate-pair exists).

Next we consider the relationship between the decisional assumption and the parent functional problem. The following fact is immediate:

Fact 2.10 *The existence of a polynomial-time algorithm for* $\mathrm{PR}[\mathbf{z}, k, t]$ *violates* $\mathrm{DPR}[\mathbf{z}, k, t]$.

Ideally, we would like to show the reverse direction as well, i.e., that any polynomial-time distinguisher for the DPR assumption implies a polynomial-time algorithm for solving the PR problem. We will clarify the issues that are pertaining to this reduction in the remaining of this section by introducing the notion of a strong location oracle and presenting a reduction to PR. A strong location oracle will be able to decide whether a given location is an error or not with an error probability that is independent of the given instance. Note that the violation of the DPR assumption does not necessarily imply the existence of a strong location oracle: this is due to the fact that PR is not randomly self-reducible. Still, the result in the remaining of the section demonstrates that local decisions, such as those employed in the DPR assumption, may imply decodability if they are instance-independent in an algorithmic sense.

For a family of PR instances $\mathcal{I}_{\mathbf{z}}$ we define the "codeword remaining redundancy" to be the ratio $(t-k)/n$. The larger the remaining redundancy is, the easier the decoding problem appears to be. For example if the codeword remaining redundancy is at least $(1-k/n)/2$ then any PR instance is fully decodable.

Definition 2.11 *A* strong location oracle $\mathcal{O}_\rho$ *for the PR problem with error* ϵ *is a TM that given any* $\langle i, \mathbf{y} \rangle$ *where* $\mathbf{y}$ *is a* $\mathrm{PR}[\mathbf{z}, k, t]$ *instance with support* $\mathbf{z}$ *and* $i \in \{1, \ldots, n\}$ *so that* $(t-k)/n \geq \rho$*, it returns* $\{0, 1\}$ *depending on whether* i *is an error-location (*$i \notin I_{\mathbf{y}}$*) or an uncorrupted location (*$i \in I_{\mathbf{y}}$*), respectively with probability* $1 - \epsilon$.

In the following lemma we show that a strong location oracle for remaining redundancy ρ can be used for implementing a decoding algorithm for the PR problem for any choice of parameters that satisfy $(t-k)/n \geq \rho$. We note that a strong location oracle would work even for instances that have ambiguous error vectors; in this case, the oracle is assumed to select one of the possible polynomial solutions (i.e., it has a preferred polynomial solution that is automatically selected).

Proposition 2.12 *Given a strong location oracle* $\mathcal{O}_\rho$ *with error* ϵ*, there exists a polynomial-time algorithm* $\mathcal{A}^{\mathcal{O}_\rho}$ *that solves the* $\mathrm{PR}[\mathbf{z}, k, t]$ *problem with* $(t-k)/n \geq \rho$ *with probability at least* $1 - n \cdot \epsilon$.

Proof. The algorithm $\mathcal{A}$ operates as follows: given $\mathbf{y} = \langle y_1, \ldots, y_n \rangle \in \mathcal{I}_{\mathbf{z},k,t}$ it submits $(\mathbf{y}, n)$ to $\mathcal{O}_{n,k,t}$. In case the answer is 1, $\mathcal{A}$ performs the following transformation $y'_i = (y_i - y_n)(z_i - z_n)^{-1}$ and forms the instance $\mathbf{y}' = \langle y'_1, \ldots, y'_{n-1} \rangle$. Note that $\mathbf{y}'$ is a PR instance with parameters $[n-1, k-1, t-1]$. On the other hand, if the answer of the location oracle is 0, $\mathcal{A}$ forms the instance $\mathbf{y}' = \langle y'_1, \ldots, y'_{n-1} \rangle$, (i.e., it simply drops the last element since it is an error location). In either case, observe that the size of the input has been reduced by 1 and we may repeat the process recursively on the instance $\mathbf{y}'$. The remaining redundancy of $\mathbf{y}'$ is either case equal to $(t-1-(k-1))/(n-1) = (t-k)/(n-1)$ and thus it holds that $(t-k)/(n-1) \geq (t-k)/n \geq \rho$. We continue the process recursively until k non-error locations of the original instance have been identified or we have reduced the parameters to satisfy the relation $t \geq (n+k)/2$ (in which case we may apply the [BW86] decoder to recover the solution). ■

3 Pseudorandomness

In this section we will present the first basic implication of the DPR assumption: the fact that PR instances are pseudorandom: i.e., computationally indistinguishable from a random set of points over $\mathbb{F}$.

In particular we will show that distinguishing instances of $\mathrm{PR}[\mathbf{z}, k, t]$ from random elements of $\mathcal{S}_n := \mathbb{F}^n$ is hard under the DPR-Assumption. We first give the definition of setindistinguishability and pseudorandomness in our setting:

Definition 3.1 *Let $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ be a family of sets parameterized by n. Two families of sets with $A_n, B_n \subseteq \mathcal{F}_n$ are* (polynomial-time, computationally) indistinguishable *if for any* PPT *predicate $\mathcal{A}$,*

$$|\ \mathsf{Pr}[\mathcal{A}(X) = 1 : X \leftarrow A_n] - \mathsf{Pr}[\mathcal{A}(X) = 1 : X \leftarrow B_n]|$$

is negligible in n. If on the other hand there is an $\mathcal{A}$ for which the probability above is non-negligible in n, we will say that $\mathcal{A}$ is a distinguisher *for A_n, B_n. A family of sets A_n is called* $\mathcal{F}$-pseudorandom *if it is indistinguishable from $\mathcal{F}_n$.*

Note that for this section we consider $B_n = \mathbb{F}^n$ and $A_n = \mathcal{I}_{\mathbf{z},k,t}$

Definition 3.2 *We define the function $\mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},k,t} = \max_{\mathcal{A}}(\mathsf{Adv}^{\mathsf{psr},\mathcal{A}}_{\mathbf{z},k,t})$ where $\mathsf{Adv}^{\mathsf{psr},\mathcal{A}}_{\mathbf{z},k,t} = |\mathsf{Pr}[\mathcal{A}(X) = 1] - \mathsf{Pr}[\mathcal{A}(Y) = 1]|$ where X is distributed according to $\mathsf{S}(\mathbf{z}, k, t)$ and Y is distributed according to U over $\mathbb{F}^n$.*

Next we present a basic probabilistic lemma that will assist in the analysis later on.

Lemma 3.3 *Let $v^{\mathsf{b}}_i, v^{\mathsf{g}}_i$ be independent samplable binary random variables for $i \in \{1, \ldots, n\}$ with means $\mu_{\mathsf{b},i}$ and $\mu_{\mathsf{g},i}$ respectively for which it holds:*

- *There exists an $i \in \{1, \ldots, n\}$ such that $|\mathsf{Pr}[v^{\mathsf{g}}_i = 1] - \mathsf{Pr}[v^{\mathsf{b}}_i = 1]| \geq \alpha$ where α is a non-negligible function in n.*

Then, for all $\epsilon > 0$, there exists a probabilistic polynomial-time TM $\mathcal{B}$ that returns an i that satisfies $|\mathsf{Pr}[v^{\mathsf{g}}_i = 1] - \mathsf{Pr}[v^{\mathsf{b}}_i = 1]| \geq \alpha/4$ with probability $1 - \epsilon$. $\mathcal{B}$ requires $\mathcal{O}(\alpha^{-2}(\log(\epsilon^{-1}) + \log n))$ samples of each of the given random variables.

Proof. Consider the following procedure $\mathcal{B}$: first it produces the samples $v_{i,j}^{\mathsf{b}}, v_{i,j}^{\mathsf{g}}$ distributed according to $v_i^{\mathsf{b}}, v_i^{\mathsf{g}}$ for $i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,m\}$ where m is a parameter to be specified later. Then, for all $i = 1,\ldots,n$, it computes the sums $s_i^{\mathsf{b}} = \sum_{j=1}^{m} v_{i,j}^{\mathsf{b}}$ and $s_i^{\mathsf{g}} = \sum_{j=1}^{m} v_{i,j}^{\mathsf{g}}$ as well as the difference $\delta_i = |s_i^{\mathsf{b}} - s_i^{\mathsf{g}}|$. Finally, $\mathcal{B}$ collects all i such that $\delta_i \geq \alpha m/2$ forming a list L and returns one of them at random. If no such i exists, i.e., the list L is empty, $\mathcal{B}$ returns $\perp$.

We turn to the analysis of $\mathcal{B}$ next. Let i_0 be the special index that is guarranteed in the statement of the theorem and satisfies $|\mathsf{Pr}[v_{i_0,1}^{\mathsf{b}} = 1] - \mathsf{Pr}[v_{i_0,1}^{\mathsf{g}} = 1]| \geq \alpha$. Using the Chernoff bound we have that $\mathsf{Pr}[|s_i^{\mathsf{g}} - \mu_{\mathsf{g},i}| \leq t_1] \geq 1 - e^{-t_1^2/2m}$ and $\mathsf{Pr}[|s_i^{\mathsf{b}} - \mu_{\mathsf{b},i}| \leq t_2] \geq 1 - e^{-t_2^2/2m}$ for any $t_1, t_2 > 0$. Using the condition we have for i_0 we obtain that $|\mu_{\mathsf{g},i_0} - \mu_{\mathsf{b},i_0}| \geq \alpha m$. We set $t_1 = t_2 = \alpha m/4$ and we obtain that with probability at least $1 - 2e^{-\alpha^2 m/2}$ it happens that the events $|s_{i_0}^{\mathsf{g}} - \mu_{\mathsf{g},i_0}| \leq \alpha m/4$ and $|s_{i_0}^{\mathsf{b}} - \mu_{\mathsf{b},i_0}| \leq \alpha m/4$ are simultaneously true. This implies that $|s_{i_0}^{\mathsf{g}} - s_{i_0}^{\mathsf{b}}| \leq \alpha m/2$ is also true so we conclude that $\mathsf{Pr}[\delta_{i_0} \geq \alpha m/2] \geq 1 - 2e^{-\alpha^2 m/2}$. It follows that with this probability the index i_0 will be one of the possible indices that $\mathcal{B}$ can return as output (and as a result the list of such possible indices is not empty conditioning on this event).

Suppose now for some index $i_1 \in \{1,\ldots,n\}$ it holds $|\mathsf{Pr}[v_{i_1,1}^{\mathsf{g}} = 1] - \mathsf{Pr}[v_{i_1,1}^{\mathsf{b}} = 1]| < \tau \cdot \alpha$ which implies that $|\mu_{\mathsf{g},i_1} - \mu_{\mathsf{b},i_1}| < \tau\alpha m$, for some $\tau \in (0,1)$. Using the Chernoff bound again we have that $\mathsf{Pr}[|s_{i_1}^{\mathsf{g}} - \mu_{\mathsf{g},i_1}| \leq t_1] \geq 1 - e^{-t_1^2/2m}$ and $\mathsf{Pr}[|s_{i_1}^{\mathsf{b}} - \mu_{\mathsf{b},i_1}| \leq t_2] \geq 1 - e^{-t_2^2/2m}$ for any $t_1, t_2 > 0$. We set $t_1 = t_2 = \tau\alpha m/2$ and we obtain that the probabilities of the events $|s_{i_1}^{\mathsf{g}} - \mu_{\mathsf{g},i_1}| \leq \tau\alpha m/2$ and $|s_{i_1}^{\mathsf{b}} - \mu_{\mathsf{g},i_1}| \leq \tau\alpha m/2$ are both at least $1 - e^{-\tau^2\alpha^2 m/8}$. Based on this and the bound we have on the means $|\mu_{\mathsf{g},i_1} - \mu_{\mathsf{b},i_1}| < \tau\alpha m$ we conclude that $\mathsf{Pr}[\delta_{i_1} < 2\tau\alpha m] \geq 1 - 2e^{-\tau^2\alpha^2 m/8}$, i.e., $\mathsf{Pr}[\delta_{i_1} \geq 2\tau\alpha m] \leq 2e^{-\tau^2\alpha^2 m/8}$. We set $\tau = 1/4$ and we obtain that $\mathsf{Pr}[\delta_{i_1} \geq \alpha m/2] \leq 2e^{-\alpha^2 m/128}$.

To complete the analysis we need to provide a lower bound for the probability to output an index that satisfies $|\mathsf{Pr}[v_{i,1}^{\mathsf{g}} = 1] - \mathsf{Pr}[v_{i,1}^{\mathsf{b}} = 1]| \geq \alpha/4$. Note that in the worst case there will be $n-1$ indices i that satisfy the condition $|\mathsf{Pr}[v_{i,1}^{\mathsf{g}} = 1] - \mathsf{Pr}[v_{i,1}^{\mathsf{b}} = 1]| < \alpha/4$ (excluding i_0). The event we are interested in is the following: none of the indices with distinguishing probability less than $\alpha/4$ belong to the list L while the index i_0 belongs to the list L. Using the above arguments we conclude that the probability we are interested in is at least $(1 - 2e^{-\alpha^2 m/128})^{n-1}(1 - 2e^{-\alpha^2 m/2}) \geq 1 - 2(n-1)e^{-\alpha^2 m/128} - 2e^{-\alpha^2 m/2} \geq 1 - e^{-\alpha^2 m/128 + \ln 2n}$ which is greater equal to $1 - \epsilon$ provided we set $m \geq 128 \cdot \alpha^{-2}(\ln(\epsilon)^{-1} + \ln 2n)$. Recall that based on the fact that α is a non-negligible function in n it holds that m is polynomial in $n, \log \epsilon^{-1}$ as required in the theorems statement. ■

We proceed to the main theorem of this section that establishes the pseudorandomness of the distribution induced by S (cf. definition 2.6) which implies that PR instances are pseudorandom under the decisional PR assumption.

Theorem 3.4 *Let $\mathcal{A}$ be a* PPT *predicate that is a polynomial-time distinguisher of the distribution $\mathsf{S}_{n,k,t}^{\mathbf{z}}$ over $\mathbb{F}^n$ and the uniform distribution U over $\mathbb{F}^n$ with distinguishing probability at least α. Then, it holds that $\alpha \leq \frac{t(n-t+3)}{|\mathbb{F}|} + t \cdot \mathsf{Adv}_{\mathbf{z}',k,t}^{\mathsf{dpr}} + 8t \cdot \mathsf{Adv}_{\mathbf{z},k,t}^{\mathsf{dpr}}$, where $\mathbf{z}' \in \mathbb{F}^{n-1}$ and is obtained from $\mathbf{z}$ by removing one of the coordinates.*

Proof. Let $\mathcal{A}$ be the distinguisher between the distributions $\mathsf{S}_{n,k,t}^{\mathbf{z}}$ and U as described in the theorem's statement. Also define the sampler S_i to denote the distribution of pairs $\langle i, \mathbf{y} \rangle$

where $\mathbf{y}$ is sampled according to S. Consider the following procedure $\mathcal{A}_1$ that operates on pairs of the form $\langle i, \mathbf{y}\rangle$ as follows: it first selects a random permutation π and then overwrites the $y_{\pi(1)}, \ldots, y_{\pi(i)}$ values of the vector $\mathbf{y}$ (provided that $i > 0$) by substituting them with i random values over $\mathbb{F}$; in this way $\mathcal{A}_1$ produces the "partially randomized" PR instance $\mathbf{y}'$. Then $\mathcal{A}_1$ simulates $\mathcal{A}$ on $\mathbf{y}'$. We will denote the operation of $\mathcal{A}_1$ as $\mathcal{A}(\mathsf{R}_i^\pi(\mathbf{y}))$ where R_i^π is the probabilistic operator that given $\mathbf{y}$, it randomizes the first (according to π) i locations of $\mathbf{y}$.

It is immediate that

$$\mathsf{Pr}[\mathcal{A}_1(\mathsf{S}_0(\mathbf{z}, k, t)) = 1] = \mathsf{Pr}[\mathcal{A}(\mathsf{S}(\mathbf{z}, k, t)) = 1]$$

as well as that

$$\mathsf{Pr}[\mathcal{A}_1(\mathsf{S}_n(\mathbf{z}, k, t)) = 1] = \mathsf{Pr}[\mathcal{A}(\mathsf{U}) = 1]$$

As a result $|\mathsf{Pr}[\mathcal{A}_1(\mathsf{S}_0(\mathbf{z}, k, t)) = 1] - \mathsf{Pr}[\mathcal{A}_1(\mathsf{S}_n(\mathbf{z}, k, t)) = 1]| \geq \alpha$ since $|\mathsf{Pr}[\mathcal{A}(\mathsf{S}(\mathbf{z}, k, t)) = 1] - \mathsf{Pr}[\mathcal{A}(\mathsf{U}) = 1]| \geq \alpha$ from the statement of the theorem. By employing the triangular inequality we obtain that there exists $i \in \{1, \ldots, n\}$ such that

$$|\mathsf{Pr}[\mathcal{A}_1(\mathsf{S}_i(\mathbf{z}, k, t)) = 1] - \mathsf{Pr}[\mathcal{A}_1(\mathsf{S}_{i-1}(\mathbf{z}, k, t)) = 1]| \geq \alpha/n$$

Below we will denote by $\mathsf{E}_{n,k,t}^{i,\pi}$ the event $\mathcal{A}(\mathsf{R}_i^\pi(\mathsf{S}(\mathbf{z}, k, t))) = 1$. Note that we don't specify $\mathbf{z}$ in the E notation as we will use the same $\mathbf{z}$ in conjunction to E in the remaining of the proof. Using this notation and the above results we obtain that :

$$\forall \pi \ \exists i \in \{1, \ldots, n\} \text{ s.t. } |\mathsf{Pr}[\mathsf{E}_{n,k,t}^{i,\pi}] - \mathsf{Pr}[\mathsf{E}_{n,k,t}^{i-1,\pi}]| \geq \alpha' \tag{1}$$

where $\alpha' = \alpha/n$.

Next, consider the event Bad_i^π to correspond to the coin tosses of the sampler $\mathsf{S}(\mathbf{z}, k, t)$ that the location $\pi(i)$ is among the error-locations. We denote by Good_i^π the negation of this event.

Claim 1. $|\mathsf{Pr}[\mathsf{E}_{n,k,t}^{i,\pi} \mid \mathsf{Bad}_i^\pi] - \mathsf{Pr}[\mathsf{E}_{n,k,t}^{i-1,\pi} \mid \mathsf{Bad}_i^\pi]| \leq 1/|\mathbb{F}|$.

Indeed, observe that in the conditional space Bad_i^π for the sampler S the $\pi(i)$-th location of the vector $\mathbf{y}$ is distributed uniformly over the set $\mathbb{F} - \{p(z_{\pi(i)})\}$ where p is the solution polynomial that is selected by the sampler. The probabilistic operator R_i^π will substitute the $\pi(i)$-th location with a random element over $\mathbb{F}$. It follows by a standard argument that the statistical distance between the two distributions is at most $1/|\mathbb{F}|$ from which the claim 1 follows.

Next we use the fact: if $|\mathsf{Pr}[E_1] - \mathsf{Pr}[E_2]| \geq \alpha$ and $|\mathsf{Pr}[E_1|B] - \mathsf{Pr}[E_2|B]| \leq \epsilon$ then it holds that $|\mathsf{Pr}[E_1|\neg B] - \mathsf{Pr}[E_2|\neg B]| \geq (\alpha - \epsilon \cdot \mathsf{Pr}[B])(\mathsf{Pr}[\neg B])^{-1}$. Applying this on claim 1 we obtain the following:

$$|\mathsf{Pr}[\mathsf{E}_{n,k,t}^{i,\pi} \mid \mathsf{Good}_i^\pi] - \mathsf{Pr}[\mathsf{E}_{n,k,t}^{i-1,\pi} \mid \mathsf{Good}_i^\pi]| \geq \alpha'' \tag{2}$$

where $\alpha'' = \frac{n}{t}(\alpha' - (1 - \frac{t}{n})|\mathbb{F}|^{-1}) = \frac{\alpha}{t} - \frac{n-t}{|\mathbb{F}|}$.

Claim 2. $\mathsf{Pr}[\mathsf{E}_{n,k,t}^{i,\pi} \mid \mathsf{Good}_i^\pi] = |\mathsf{Pr}[\mathsf{E}_{n,k,t-1}^{i,\pi} \mid \mathsf{Bad}_i^\pi]$.

The validity of the second claim can be established by directly corresponding the random coins of event $\mathsf{E}_{n,k,t}^{i,\pi}$ in the conditional space Good_i^π to the random coins of event $\mathsf{E}_{n,k,t-1}^{i,\pi}$ in the conditional space Bad_i^π. The event $\mathsf{E}_{n,k,t}^{i,\pi}$ can be thought of containing tuples of the form

$\langle I_L, p_L, \mathbf{e}_L, \vec{r}_L\rangle$ so that I_L is a subset of $\{1,\ldots,n\}$ that necessarily includes i, $p_L \in \mathbb{F}[x]$ with degree$(p_L) < k$, $\mathbf{e}_L$ is a (n,t)-error-vector that is zero in I_L and finally $\vec{r}_L$ is a random vector of $\mathbb{F}^i$ that specifies the coins of the probabilistic operator R_i^π. On the other hand, the event $\mathsf{E}_{n,k,t-1}^{i,\pi}$ in the conditional space Bad_i^π can be thought of containing tuples of the form $\langle I_R, p_R, \mathbf{e}_R, \vec{r}_R\rangle$ where I_R is a subset of $\{1,\ldots,n\}$ with cardinality $t-1$, $p_R \in \mathbb{F}[x]$ with degree$(p_R) < k$, $\mathbf{e}_R$ is a $(n,t-1)$-error-vector that is zero in I_R and $\vec{r}$ is a random vector of $\mathbb{F}^i$ that defines the coins of the probabilistic operator R_i^π. Consider the following correspondence: given a tuple $\langle I_L, p_L, \mathbf{e}_L, \vec{r}_L\rangle$ we define $\langle I_R, p_R, \mathbf{e}_R, \vec{r}_R\rangle$ as follows: $I_R = I_L - \{\pi(i)\}, p_R = p_L, \vec{r}_R = \vec{r}_L$ and also we set $(\mathbf{e}_R)_j = (\mathbf{e}_L)_j$ for all $j \neq \pi(i)$ (note that $(\mathbf{e}_L)_{\pi(i)} = 0$ since $\pi(i)$ is not an error location). Finally we select $(\mathbf{e}_R)_{\pi(i)}$ at random from $\mathbb{F} - \{p_R(z_{\pi(i)})\}$. We remark that the choice of $(e_R)_{\pi(i)}$ does not affect the outcome of the experiment since it substituted with the same random value in both cases. It follows that for every tuple of $\mathsf{E}_{n,k,t}^{i,\pi}$ in the conditional space Good_i^π we have a correspondence of the same number of tuples of $\mathsf{E}_{n,k,t-1}^{i-1,\pi}$ in the conditional space Bad_i^π. Based on this the statement of the claim follows.

Claim 3. $|\mathsf{Pr}[\mathsf{E}_{n,k,t-1}^{i,\pi} \mid \mathsf{Bad}_i^\pi] - |\mathsf{Pr}[\mathsf{E}_{n,k,t}^{i-1,\pi} \mid \mathsf{Bad}_i^\pi]| \leq \mathsf{Adv}_{\mathbf{z}',k,t}^{\mathsf{dpr}} + 3/|\mathbb{F}|$.

Recall that the event $\mathsf{E}_{n,k,t}^{i,\pi}$ is defined as $\mathcal{A}(\mathsf{R}_i^\pi(\mathsf{S}(\mathbf{z},k,t))) = 1$. We will argue that the two probability ensembles $\mathsf{R}_i^\pi(\mathsf{S}(\mathbf{z},k,t-1))$ and $\mathsf{R}_{i-1}^\pi(\mathsf{S}(\mathbf{z},k,t))$ are computationally indistinguishable when considered over the conditional probability spaces based on the event Bad_i^π. Suppose that $\mathcal{B}$ is any PPT distinguisher between the two ensembles. We define next a PPT distinguisher $\mathcal{B}'$ for the DPR$[\mathbf{z}',k,t]$ that operates as follows: $\mathcal{B}'$ given the challenge $\langle j, \mathbf{y}\rangle$ over the support set $\mathbf{z}' = \langle z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n\rangle$, $\mathcal{B}'$ first randomizes the value $(\mathbf{y})_j$; then it parses $\mathbf{y}$ as $\langle y_1,\ldots,y_{\pi(i)-1},y_{\pi(i)+1},\ldots y_n\rangle$ and injects a random value of $\mathbb{F}$ at location $\pi(i)$; finally it selects $y'_{\pi(1)},\ldots,y'_{\pi(i-1)}$ from $\mathbb{F}$ and overwrites the corresponding $i-1$ locations of $\mathbf{y}$. The resulting vector $\mathbf{y}_{\mathsf{new}}$ is of length n. $\mathcal{B}'$ terminates by simulating $\mathcal{B}$ on $\mathbf{y}_{\mathsf{new}}$ and returning the output that $\mathcal{B}$ returns.

Suppose that the DPR$[\mathbf{z}',k,t]$ challenge $\langle j,\mathbf{y}\rangle$ was drawn according to the $\mathsf{S}^{\mathsf{bad}}(\mathbf{z}',k,t)$ sampler. This means that the vector $\mathbf{y}$ with the j-th location randomized is at a statistical distance $1/|\mathbb{F}|$ from $\mathsf{S}(\mathbf{z}',k,t)$ and after the injection of the random $\pi(i)$-th location value it will be at a statistical distance $2/|\mathbb{F}|$ from $\mathsf{S}(\mathbf{z},k,t)$ in the conditional probability space based on Bad_i^π. On the other hand, in the case that the DPR$[\mathbf{z}',k,t]$ challenge $\langle j,\mathbf{y}\rangle$ was drawn according to the $\mathsf{S}^{\mathsf{good}}(\mathbf{z}',k,t)$ sampler we would have the following: the vector $\mathbf{y}$ with the j-th location randomized is at a statistical distance $1/|\mathbb{F}|$ from $\mathsf{S}(\mathbf{z}',k,t-1)$; it follows that, after injecting the $\pi(i)$-th location element and randomizing the $i-1$ locations according to π, the resulting vector $\mathbf{y}_{\mathsf{new}}$ is at a distance $1/|\mathbb{F}|$ from the ensemble $\mathsf{R}_i^\pi(\mathsf{S}(\mathbf{z},k,t-1))$. From these facts the statement of claim 3 follows.

By applying the results of claim 2 and 3 to the inequality 2 we obtain the following :

$$|\mathsf{Pr}[\mathsf{E}_{n,k,t}^{i-1,\pi} \mid \mathsf{Bad}_i^\pi] - \mathsf{Pr}[\mathsf{E}_{n,k,t}^{i-1,\pi} \mid \mathsf{Good}_i^\pi]| \geq \alpha''' \tag{3}$$

where $\alpha''' = \frac{\alpha}{t} - \frac{n-t+3}{|\mathbb{F}|} - \mathsf{Adv}_{\mathbf{z}',k,t}^{\mathsf{dpr}}$. Using the definition of the event $\mathsf{E}_{n,k,t}^{i-1,\pi}$ we rewrite equation 3 as follows:

$$|\mathsf{Pr}[\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathsf{S}(\mathbf{z},k,t))) = 1 \mid \mathsf{Bad}_i^\pi] - \mathsf{Pr}[\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathsf{S}(\mathbf{z},k,t))) = 1 \mid \mathsf{Good}_i^\pi]| \geq \alpha''' \tag{4}$$

Where i is some index in $\{1,\ldots,n\}$ that while it is unknown, its existence is guaranteed

from equation 1. Next we observe that we can simulate the behavior of the sampler S in the conditional probability spaces Bad_i^π and Good_i^π. In particular this can be done easily by the samplers $\mathsf{S}^{\mathsf{Bad}_i^\pi}$ and $\mathsf{S}^{\mathsf{Good}_i^\pi}$ that operate exactly as S with the exception the selection of the set of indices I that is done as follows: for the case of $\mathsf{S}^{\mathsf{Good}_i^\pi}$ a random subset $I \subseteq \{1,\ldots,n\} - \{\pi(i)\}$ is selected that has cardinality $t-1$ and then the element $\pi(i)$ is added to it; on the other hand, for the case of $\mathsf{S}^{\mathsf{Bad}_i^\pi}$ a random subset $I \subseteq \{1,\ldots,n\} - \{\pi(i)\}$ is selected with cardinality t. Based on this it follows that we can rewrite equation 4 in this way:

$$|\mathsf{Pr}[\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathsf{S}^{\mathsf{Bad}_i^\pi}(\mathbf{z},k,t))) = 1] - \mathsf{Pr}[\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathsf{S}^{\mathsf{Good}_i^\pi}(\mathbf{z},k,t))) = 1]| \geq \alpha''' \tag{5}$$

Observe that if we define $u_{i,j}^{\mathsf{b}}(\pi)$ to be equal to $\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathsf{S}^{\mathsf{Bad}_i^\pi}(\mathbf{z},k,t)))$ and $u_{i,j}^{\mathsf{g}}(\pi)$ to be equal to $\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathsf{S}^{\mathsf{Good}_i^\pi}(\mathbf{z},k,t)))$, the conditions of lemma 3.3 are satisfied (for any permutation π). Armed with this observation we describe the following PPT procedure $\mathcal{A}_2$ that acts as a $\mathrm{DPR}[\mathbf{z},k,t]$ distinguisher over the support elements $\mathbf{z}$.

$\mathcal{A}_2$ on input $\langle j, \mathbf{y}\rangle$ operates as follows. First $\mathcal{A}_2$ executes the following loop that depends only on the input j:

1. Choose a random permutation π over $\{1,\ldots,n\}$.
2. Execute the procedure $\mathcal{B}$ of lemma 3.3 to find the special index i for which equation 5 holds using the random variables $u_{i,j}^{\mathsf{b}}(\pi), u_{i,j}^{\mathsf{g}}(\pi)$ (note that the execution of the lemma is based on parameters $[\mathbf{z}^\pi, k, t]$ and provides an i that satisfies equation 5 with success $\alpha'''/4$).
3. If $\pi(i) = j$ stop and return (i,π) otherwise repeat from step 1.

Recall that the procedure $\mathcal{B}$ of lemma 3.3 will return an index that satisfies equation 5 with threshold $\alpha'''/4$ and will succeed with probability $1-\epsilon_1$ where ϵ_1 is a parameter we will specify. The procedure $\mathcal{B}$ runs in time polynomial in $n + \ln \epsilon_1^{-1} + (\alpha''')^{-2}$.

Regarding the number of repetitions that are required for exiting the above loop at step 3 observe the following: the value j is independent of the determination of i, π. It follows that with probability $1/n$ the loop will terminate at step 3 after the first repetition. After $\ln \epsilon_2^{-1} \cdot n$ repetitions of the loop we conclude that the probability of failing all $\ln \epsilon_2^{-1} \cdot n$ times is $(1-1/n)^{\ln \epsilon_2^{-1} \cdot n} \leq \epsilon_2$.

It follows that $\mathcal{A}_2$ will terminate step 3 successfully and with probability at least $1-\epsilon_1-\epsilon_2 = 1-\epsilon$, it will possess at this stage an index i and a permutation π so that the equation 5 exhibits a spike at location i for parameters $\mathbf{z}, k, t$. Moreover it holds that $\pi(i) = j$.

After step 3 terminates, $\mathcal{A}_2$ simulates $\mathcal{A}(\mathsf{R}_{i-1}^\pi(\mathbf{y}))$ with parameters $\mathbf{z}, k, t$ and returns the output that $\mathcal{A}$ returns.

Suppose that $\langle j, \mathbf{y}\rangle$ is distributed according to $\mathsf{S}^{\mathsf{bad}}(\mathbf{z},k,t)$. It follows that $\mathbf{y}$ will be distributed according to $\mathsf{S}^{\mathsf{Bad}_i^\pi}(\mathbf{z}^\pi,k,t)$. With a similar argument we obtain that if $\langle j, \mathbf{y}\rangle$ is distributed according to $\mathsf{S}^{\mathsf{good}}(\mathbf{z},k,t)$, the vector $\mathbf{y}$ will be distributed according to $\mathsf{S}^{\mathsf{Good}_i^\pi}(\mathbf{z},k,t)$.

Next we use the fact that if $|\mathsf{Pr}[E_1|N] - \mathsf{Pr}[E_2|N]| \geq \alpha$ and $\mathsf{Pr}[\neg N] \leq \epsilon$ then $|\mathsf{Pr}[E_1] - \mathrm{Pr}[E_2]| \geq (1-\epsilon)\alpha - \epsilon$. Based on the above we conclude that $\mathcal{A}_2$ is a distinguisher for $\mathrm{DPR}[\mathbf{z},k,t]$ with distinguishing probability at least $(1-\epsilon)\alpha'''/4 - \epsilon$. It follows that $(1-\epsilon)\alpha'''/4 - \epsilon \leq \mathsf{Adv}_{\mathbf{z},k,t}^{\mathsf{dpr}}$ and using that $\alpha''' = \frac{\alpha}{t} - \frac{n-t+3}{|\mathbb{F}|} - \mathsf{Adv}_{\mathbf{z}',k,t}^{\mathsf{dpr}}$ we conclude that $\alpha \leq t(n-t+3)/|\mathbb{F}| + t \cdot \mathsf{Adv}_{\mathbf{z}',k,t}^{\mathsf{dpr}} + 8t \cdot \mathsf{Adv}_{\mathbf{z},k,t}^{\mathsf{dpr}}$ by setting $\epsilon \leq \mathsf{Adv}_{\mathbf{z},k,t}^{\mathsf{dpr}}$ (note that if $\mathsf{Adv}_{\mathbf{z},k,t}^{\mathsf{dpr}}$ is exponentially small in a parameter s then the parameters $\ln \epsilon_1^{-1}$ and $\ln \epsilon_2^{-1}$ with $\epsilon_1 + \epsilon_2 = \epsilon$ will be polynomial in s). ■

Corollary 3.5 *It holds that* $\mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},k,t} \leq \frac{t(n-t+3)}{|\mathbb{F}|} + t \cdot \mathsf{Adv}^{\mathsf{dpr}}_{\mathbf{z}',k,t} + 8t \cdot \mathsf{Adv}^{\mathsf{dpr}}_{\mathbf{z},k,t}$.

4 Hardness of Recovering Partial Information of Polynomial Values

In this section we will show that $\mathrm{PR}[\mathbf{z}, k, t]$ "leaks no partial information" about any specific polynomial value under the DPR-Assumption even if these values are distributed according to an arbitrary probability distribution. This result suggests that the decisional PR assumption enables a form of semantic security for specific points of the polynomial curve that it hides.

We introduce first the following sampler that restricts the PR sampler of definition 2.6.

Definition 4.1 *The sampler* S^w *over* $\mathbb{F}^n$ *is defined as follows:* S^w *given* $\langle \mathbf{z}, k, t\rangle$ *such that* $\{0, \ldots, u-1\} \cap \{z_1, \ldots, z_n\} = \emptyset$ *and and* $w = \langle w_0, \ldots, w_{u-1}\rangle \in \mathbb{F}^u$*, it first interpolates a polynomial* $p \in \mathbb{F}[x]$ *such that (1)* $p(i) = w_i$ *for* $i = 0, \ldots, u-1$ *and (2)* $p(i)$ *is uniformly distributed for* $i = u, \ldots, k$*. Finally it selects* $\mathbf{e}$*, a random* (n,t)*-error-vector and returns* $\mathbf{p}+\mathbf{e}$*.*

We note that S^w induces a distribution over $\mathbb{F}^n$ that would be identical to that of the sampler S of definition 2.6 if w is a uniform random variable over $\mathbb{F}^u$. We denote the probability distribution induced by S^w over $\mathbb{F}^n$ as $\mathsf{S}^w_{\mathbf{z},k,t}$. For the remaining of the section we will assume that $\{z_1, \ldots, z_n\} \cap \{0, 1, \ldots, u-1\}$ where u will be clear from the context.

We model next what it means for a PR instance $\mathbf{y}$ to leak no partial information about a certain portion of its polynomial solution $p_{\mathbf{y}}$.

Definition 4.2 *We say that* $\mathrm{PR}[\mathbf{z}, k, t]$ leaks no partial information *for* u *points, if for all samplable polynomial time distributions* $\mathcal{D}$ *over* $\mathbb{F}^u$ *with* $u < k$*, for any* $g : \mathbb{F}^u \to R$ *and for any* PPT $\mathcal{A}$ *it holds that there exists a* PPT Sim *that satisfies the following :*

$$|\mathsf{Pr}[\mathcal{A}(y) = g(w) : \mathbf{y} \leftarrow \mathsf{S}^w_{\mathbf{z},k,t}, w \leftarrow \mathcal{D}] - \mathsf{Pr}[\mathsf{Sim}^{\mathcal{A}}(1^n) = g(w) : w \leftarrow \mathcal{D}]| = \mathsf{negl}(\lambda)$$

The rationale behind the above definition is that for any distribution $\mathcal{D}$ that specifies a certain portion of the polynomial p, and for any computable function g that an adversary $\mathcal{A}$ wishes to compute over the polynomial solution of the instance, it holds that it is possible to simulate the output that $\mathcal{A}$ obtains given $\mathbf{y}$ without having access to $\mathbf{y}$ (i.e., releasing $\mathbf{y}$ to the adversary does not provide any additional information that can be used to evaluate the function g over the specified portion of the instance's polynomial solution).

We show next that under the pseudorandomness of PR instances it holds that PR leaks no partial information.

Theorem 4.3 *There exists a* PPT Sim *such that for any polynomial time samplable distribution* $\mathcal{D}$ *over* $\mathbb{F}^u$*, any* $g : \mathbb{F}^u \to R$ *and any* PPT $\mathcal{A}$ *it holds that*

$$|\mathsf{Pr}[\mathcal{A}(\mathbf{y}) = g(w) : \mathbf{y} \leftarrow \mathsf{S}^w_{\mathbf{z},k,t}, w \leftarrow \mathcal{D}] - \mathsf{Pr}[\mathsf{Sim}^{\mathcal{A}}(1^n) = g(w) : w \leftarrow \mathcal{D}]| \leq \mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},k-u,t}$$

Proof. Consider the following PPT Sim that operates using an oracle call to a procedure $\mathcal{A}$ (the adversary):

1. Sample $\mathbf{y}$ from $\mathbb{F}^n$ according to U.
2. Call $\mathcal{A}$ on input $\mathbf{y}$ to obtain out.
3. return out.

We will prove that $\mathsf{Sim}^{\mathcal{A}}$ as defined above satisfies the statement of the theorem. To see this, consider the following distinguisher machine $\mathcal{B}$ that operates on a given input $\mathbf{y} \in \mathbb{F}^n$:

1. Sample $\langle w_0, \ldots, w_{u-1} \rangle \stackrel{\mathcal{D}}{\leftarrow} \mathbb{F}^u$;
2. Interpolate $q \in \mathbb{F}[x]$ such that $q(i) = w_i$ for $i = 0, \ldots, u-1$.
3. Compute $\mathbf{y}^{\mathsf{new}}$ as follows $(\mathbf{y}^{\mathsf{new}})_i = (\mathbf{y})_i \cdot \prod_{\ell=0}^{u-1}(z_i - \ell) + q(z_i)$.
4. Call $\mathcal{A}$ on input $\mathbf{y}^{\mathsf{new}}$ to obtain out.
5. If $\mathsf{out} = g(u)$ then output 1

Suppose now that the above procedure $\mathcal{B}$ is given input that is distributed according to the uniform distribution U over $\mathbb{F}^n$. It is easy to see that in this case the tuple $\mathbf{y}^{\mathsf{new}}$ is distributed also according to U over $\mathbb{F}^n$ (as it is merely a point-wise linear transformation of $\mathbf{y}$) and as a result calling $\mathcal{A}$ on $\mathbf{y}^{\mathsf{new}}$ is precisely the operation of $\mathsf{Sim}^{\mathcal{A}}$. It follows that the event $\mathsf{Sim}^{\mathcal{A}}(1^n) = g(w)$ would be equal to the event $\mathcal{B}(\mathbf{y}) = 1$ when $\mathbf{y}$ is distributed according to U.

Suppose next that the procedure $\mathcal{B}$ is given input that is distributed according to $\mathsf{S}_{\mathbf{z},k-u,t}$. Given such vector the reader can verify that $\mathbf{y}^{\mathsf{new}}$ is distributed according to $\mathsf{S}^{w}_{\mathbf{z},k,t}$ where w is distributed according to $\mathcal{D}$. Based on this it follows that the event $\mathcal{A}(\mathbf{y}) = g(w)$ would be equal to the event $\mathcal{B}(\mathbf{y}) = 1$ when $\mathbf{y}$ is distributed according to $\mathsf{S}_{\mathbf{z},k-u,t}$. The statement of the theorem follows. ■

The following corollary is immediate based on the results of the previous section:

Corollary 4.4 *Under the* $\mathrm{DPR}[n, k-u, t]$ *assumption it holds that* $\mathrm{PR}[\mathbf{z}, k, t]$ *leaks no partial information for* u *points.*

In the rest of the section we present special cases of the above theorem that are common in the cryptographic setting. Let us assume that the distribution $\mathcal{D}$ is uniform and $u = 1$. Let $g : \mathbb{F} \to R$ be a collection of poly-time computable functions defined over any $\mathbb{F}$ (for simplicity we write g for each member of this collection). Define $\mathbb{F}_a = \{u \mid g(u) = a; u \in \mathbb{F}\}$ for any $a \in R$. We say that g is *balanced* if for all $a \in R$ and all polynomials q it holds that $\mid \frac{|\mathbb{F}_a|}{|\mathbb{F}|} - \frac{1}{|R|} \mid < \frac{1}{q(\log |\mathbb{F}|)}$ (for sufficiently large $|\mathbb{F}|$). The balanced property means that any image under g corresponds to roughly the same number of pre-images. This is a very general condition that applies to individual bits of randomly chosen elements of $\mathbb{F}$ as well as to various length bit-sequences of randomly chosen elements of $\mathbb{F}$.

Naturally, guessing an unknown value of a balanced function with a uniformly distributed pre-image cannot be done with probability significantly greater than $1/|R|$:

Fact 4.5 *Let* $g : \mathbb{F} \to R$ *be balanced, poly-time computable. Then, for any* PPT $\mathcal{A}'$, *if* $\alpha'(n) := \mathsf{Pr}[\mathcal{A}'(r') = g(u) : r' \leftarrow \mathcal{R}'; u \leftarrow \mathbb{F}]$ *it holds that* $\mid \alpha'(n) - \frac{1}{|R|} \mid$ *is negligible in* $\log |\mathbb{F}|$.

Proof. Let $\mathcal{R}'_a := \{r' \mid \mathcal{A}'(r') = a\}$ for any $a \in R$. Note that it holds that $\cup_{a \in R} \mathcal{R}'_a = \mathcal{R}'$. Let q be any polynomial; now because g is balanced:

$$\alpha'(n) = \frac{\sum_{a \in R} |\mathbb{F}_a||\mathcal{R}'_a|}{|\mathbb{F}||\mathcal{R}'|} < \frac{\sum_{a \in R} |\mathcal{R}'_a|}{|\mathcal{R}'|}\left(\frac{1}{|R|} + \frac{1}{q(\log |\mathbb{F}|)}\right) = \frac{1}{|R|} + \frac{1}{q(\log |\mathbb{F}|)}$$

and

$$\alpha'(n) = \frac{\sum_{a\in R} |\mathbb{F}_a||\mathcal{R}'_a|}{|\mathbb{F}||\mathcal{R}'|} > \frac{\sum_{a\in R} |\mathcal{R}'_a|}{|\mathcal{R}'|}\left(\frac{1}{|R|} - \frac{1}{q(\log|\mathbb{F}|)}\right) = \frac{1}{|R|} - \frac{1}{q(\log|\mathbb{F}|)}$$

consequently $|\alpha'(n) - \frac{1}{|R|}|$ is negligible in $\log|\mathbb{F}|$. ■

The corollary of fact 4.5 and theorem 4.3 is the following:

Corollary 4.6 *For any balanced $g : \mathbb{F} \to R$, the success of any PPT $\mathcal{A}$ that given $\mathbf{y} \in \mathcal{I}_{\mathbf{z},k,t}$, computes the value $g(p_{\mathbf{y}}(0))$ is only by a negligible fraction different than $1/|R|$ unless the* $\mathrm{DPR}[n, k-1, t]$ *assumption fails.*

More specifically we can give the following examples of balanced predicates/functions that are hard to compute given a $\mathrm{PR}[\mathbf{z}, k, t]$-instance:

Proposition 4.7 *The following problems are hard under the* $\mathrm{DPR}[n, k-1, t]$*:*

1. *Let* $\mathrm{BIT}_l(a)$ *denote the l-th least significant bit of $a \in \mathbb{F}$. Given $\mathbf{y} \in \mathcal{I}_{\mathbf{z},k,t}$ predict* $\mathrm{BIT}_l(p_{\mathbf{y}}(0))$ *with non-negligible advantage where l represents any bit, except the* $\log\log|\mathbb{F}|$ *most significant.*
2. *Let* $\mathrm{BITS}_l(a)$ *denote the sequence of the l least significant bits of $a \in \mathbb{F}$. Given $\mathbf{y} \in \mathcal{I}_{\mathbf{z},k,t}$ predict* $\mathrm{BITS}_l(p_{\mathbf{y}}(0))$ *with probability $\frac{1}{2^l} + \alpha(n)$ where $\alpha(n)$ is non-negligible.*
3. *Let* $\mathrm{QR}(a)$ *be 1 iff $a \in \mathbb{F}$ is a quadratic residue. Given $\mathbf{y} \in \mathcal{I}_{n,k,t}$ predict* $\mathrm{QR}(p_{\mathbf{y}}(0))$ *with non-negligible advantage.*

Proof. Due to the corollary 4.6 we only need to show that the functions given are balanced.

(1) Let H_v denote the number of elements of $\mathbb{F}$ that their l-th LSB is v (where $v \in \{0,1\}$). We want to show that $\frac{|H_0|-|H_1|}{|\mathbb{F}|}$ is negligible in $\log|\mathbb{F}|$. Let $f := |\mathbb{F}| \mathrm{mod} 2^l$. If is easy to see that $|H_0|-|H_1| = f$ if $f \leq 2^{l-1}$ and that $|H_0|-|H_1| = 2^l - f$ if $f > 2^{l-1}$. In any case, we would like to show that $\frac{2^{l-1}}{|\mathbb{F}|}$ is negligible in $\log|\mathbb{F}|$, which is easy to establish unless $l \geq \log|\mathbb{F}| - \log\log|\mathbb{F}|$.

(2) For any bitstring $b \in \{0,1\}^l$ (where $l = 1, \ldots, \lfloor\log|\mathbb{F}|\rfloor$) it holds that H_b is either (a) $\lfloor\frac{|\mathbb{F}|}{2^l}\rfloor$ or (b) $\lfloor\frac{|\mathbb{F}|}{2^l}\rfloor + 1$. Case (a): $|\frac{|H_b|}{|\mathbb{F}|} - \frac{1}{|H|}| = |\frac{\lfloor|\mathbb{F}|/2^l\rfloor}{|\mathbb{F}|} - \frac{1}{2^l}|$ which is easy to see that is negligible in $\log|\mathbb{F}|$. Case (b) is similar.

(3) Straightforward as we assume that $\mathbb{F}$ is a field of prime order. ■

We note that the exclusion of the $\log\log|\mathbb{F}|$ most significant bits from the item (1) above is independent of our treatment as depending on the order of the field they may be easy to guess, and as a result BIT_l might not be balanced. Note that if the finite field is chosen appropriately all bits of $p_{\mathbf{y}}(0)$ will be hard: e.g. if we restrict to finite fields $\mathbb{F}$ such that there is a $c \in \mathbb{N}$: $|\mathbb{F}| - 2^{\lfloor\log|\mathbb{F}|\rfloor} \leq (\log|\mathbb{F}|)^c$ then all bits will be hard (e.g. a field of numbers modulo a Mersenne prime):

Corollary 4.8 *Under the above selection of $\mathbb{F}$ and the* $\mathrm{DPR}[n, k-1, t]$*, predicting any bit in a point z_0 of the graph of the solution polynomial of a* $\mathrm{PR}[\mathbf{z}, k, t]$ *instance is hard (where $z_0 \notin \{z_1, \ldots, z_n\}$).*

Proof. The proof is immediate from proposition 4.7 and the observation that all results of this section would hold also for choices of $z_0 \neq 0$. ■

5 Applications

5.1 A Pseudorandom Extender

A pseudorandom extender is a polynomial-time machine Ext that given input from $\{0,1\}^v$, it returns output to $\{0,1\}^{v+s}$ such that if the input is a uniformly random variable U it holds that $\mathsf{Ext}(\mathsf{U})$ is indistinguishable from the uniform random variable over $\{0,1\}^{v+s}$. More formally,

Definition 5.1 *The mapping* $\mathsf{Ext} : \{0,1\}^v \rightarrow \{0,1\}^{v+s}$ *with* $v, s \in \mathbb{N}$ *and* $s > 0$ *is an* (s,ϵ)*-pseudorandom-extender if it holds that for all* PPT $\mathcal{A}$*:* $|\mathsf{Pr}[\mathcal{A}(\mathsf{U}_{v+s}) = 1] - \mathsf{Pr}[\mathcal{A}(\mathsf{Ext}(\mathsf{U}_v)) = 1]| \leq \epsilon$*, where* $\mathsf{U}_v, \mathsf{U}_{v+s}$ *are uniformly random variables over* $\{0,1\}^v, \{0,1\}^{v+s}$ *respectively.*

Let $SS_{n,t}$ to be the set of all subsets of $\{1,\ldots,n\}$ that are of size t. We define $\mathsf{rank} : \{1,\ldots,\binom{n}{t}\} \rightarrow SS_{n,t}$ a 1-1 and onto function that enumerates all subsets from $\{1,\ldots,n\}$ that are of size t (this is a subset enumerator that can be efficiently implemented). Also let $\mathsf{bin}(x)$ denote the integer representation of the string x plus one.

Suppose next that we want to sample a subset uniformly distributed among all subsets of size t. We define $\tilde{l}_1 = \log\binom{n}{t}, \tilde{l}_2 = \log|\mathbb{F}|$ $l_1 = \lfloor\tilde{l}_1\rfloor, l_2 = \lfloor\tilde{l}_2\rfloor$ and ϵ_i as $\epsilon_i = 1 - 2^{l_i - \tilde{l}_i}$ for $i = 1, 2$. If U_{l_1} is the uniform distribution over $\{0,1\}^{l_1}$ we have the following:

Lemma 5.2 *The statistical distance of* $\mathsf{rank}(\mathsf{bin}(\mathsf{U}_{l_1}))$ *from the uniform distribution* $SS_{n,t}$ *is less or equal to* ϵ_1*.*

Proof. Let $A = \#SS_{n,t}$. The statistical distance of the two distributions equals $(2^{l_1}(1/2^{l_1} - 1/A) + (A - 2^{l_1})/A)/2 = 1 - 2^{l_1}/\binom{n}{t}$ from which the statement of the lemma follows. ■

We next define the following pseudorandom extender Ext that operates on a bitstring input. The main idea is to use the seed to fix the random coins used by the sampler of definition 2.6. Appropriate truncations are made to preserve the domain and range of the extender over bitstrings.

1. Input : a string $x \in \{0,1\}^v$.
2. Split v as $v = (n-k+t)l_2 + l_1$ and $x = x_1||\ldots||x_{n-k+t}||x_0$ where x_0 is of length l_1 and each $x_1,\ldots x_{n-k+t}$ is of length l_2.
3. Seed the sampler S of definition 2.6 with x to obtain a $\mathrm{PR}[\mathbf{z}, k, t]$ instance $\mathbf{y}$ such that $I_{\mathbf{y}} = \mathsf{rank}(\mathsf{bin}(x_0))$ and the error locations of $\mathbf{y}$ are set to $x_1,\ldots,x_{n-k+t}$.
4. Parse $\mathbf{y}$ as $\langle y_1,\ldots,y_n\rangle$ and return the sequence of bits $\langle \mathrm{BITS}_{l_2}(y_1),\ldots \mathrm{BITS}_{l_2}(y_n)\rangle$.

Note that given an input from $\{0,1\}^v$ the mapping Ext above returns a bistring that has length $v + (t-k)l_2 - l_1$. Under the assumption that the selection of $n, k, t, |\mathbb{F}|$ satisfies $(t-k)l_2 > l_1$, we have that Ext extends its input bitstring to a larger bitstring; note that selecting the parameters so that $(t-k)l_2 > l_1$ is always possible by selecting the field size l_2 to be appropriately large.

Theorem 5.3 *The mapping* $\mathsf{Ext} : \{0,1\}^v \rightarrow \{0,1\}^{v+s}$ *defined above is a* (s,ϵ)*-pseudorandom extender where* $v = (n-k+t)l_2 + l_1$ *satisfies that* $s = (t-k)l_2 - l_1 > 0$ *and* $\epsilon \leq \mathsf{Adv}^{\mathsf{psr}}_{n,k,t} + (2n - t + k)\epsilon_2 + \epsilon_1$*.*

Proof. First observe that the seeding of the sampler S that is performed within Ext does not follow the uniform distribution. Based on lemma 5.2 we have that the statistical distance of the seed random variable from the uniform distribution over a vector of $\mathbb{F}^{n-k+t}$ and a random subset of size t is at most $(n-t+k)\epsilon_2+\epsilon_1$. Moreover the statistical distance of a random $\mathbb{F}^n$ tuple from a random string in $\{0,1\}^{v+s}$ is at most $n\epsilon_2$. The statement of the theorem follows. ■

Note that not all choices of parameters for PR will make ϵ_1, ϵ_2 small even if PR can be assumed hard for such parameters. In particular, to minimize ϵ_1, ϵ_2 one should choose $|\mathbb{F}|, n$ close to powers of 2, e.g., $|\mathbb{F}|$ can be selected to be a Mersenne prime.

We remark that given a pseudorandom extender one can derive a pseudorandom number generator in an iterative fashion, cf. [Gol01].

5.2 Semantically Secure Oblivious Polynomial Evaluation

Oblivious polynomial evaluation (OPE) is a two-party protocol where player A wants to compute $P(\alpha)$ for some (secret) $\alpha \in \mathbb{F}$ of her choice, where $P \in \mathbb{F}[x]$, of degree d_P, is the secret input of player B. OPE was introduced by Naor and Pinkas in [NP99, NP06]. A way to implement OPE based on the polynomial reconstruction problem and a t-out-of-n oblivious transfer ([NP99]) is as follows: Player A, prepares a random instance $\mathbf{y} = \mathbf{p} + \mathbf{e}$ of $\mathrm{PR}[\mathbf{z}, k, t]$, so that $p_{\mathbf{y}}(0) = \alpha$, and sends it to B. Player B, parses $\mathbf{y}$ as $\langle y_1, \ldots, y_n \rangle$ and computes $Q(z_i, y_i)$ for all $i = 1, \ldots, n$, where $Q(x, y) := P(y) + Q'(x, y)$, with Q' a random polynomial of degrees d, d_P such that $Q'(0, y) = 0$. Using a t-out-of-n oblivious transfer, player A obtains t values $Q(z_i, y_i)$ that correspond to the indices of the non-error locations within $\mathbf{y}$. The parameter t is set to $d + d_P(k-1) + 1$ so that player A can interpolate $Q(x, p_{\mathbf{y}}(x))$ and as a result compute the value $Q(0, p_{\mathbf{y}}(0)) = P(\alpha)$.

Obviously the security of player A depends on the hardness of the following problem: given $\mathbf{y}$ distributed according to $\mathsf{S}^{\alpha}_{n,k,t}$, extract some information about $\alpha = p_{\mathbf{y}}(0)$. In [NP99] the security of a variant of the protocol above was claimed under a (rather strong) pseudorandomness assumption: namely that $p_{\mathbf{y}}(0)$ is pseudorandom to any poly-time observer given the transcript of the protocol obtained by player B. The type of security that we want for an oblivious polynomial evaluation protocol can be defined as follows:

Definition 5.4 *Player A is semantically secure in an OPE protocol if for any interactive* PPT *adversary $\mathcal{A}$ it holds that there exists an interactive* PPT Sim *so that for any polynomial-time samplable distribution $\mathcal{D}$ over $\mathbb{F}$ and poly-time computable $g : \mathbb{F} \to R$ it holds that if A's secret input α is distributed according to $\mathcal{D}$, we have the following :*

$$|\mathsf{Pr}[\mathcal{A}^{A(\alpha)}(1^n) = g(\alpha) : \alpha \leftarrow \mathcal{D}] - \mathsf{Pr}[\mathsf{Sim}^{\mathcal{A}}(1^n) = g(\alpha) : \alpha \leftarrow \mathcal{D}]| = \mathsf{negl}(\lambda)$$

Theorem 5.5 *Under the* $\mathrm{DPR}[n, k-1, t]$ *assumption, player A is semantically secure in the OPE protocol presented above (assuming an ideal implementation of t-out-of-n OT).*

Proof. The proof follows immediately from corollary 4.4 setting $u = 1$: $\mathrm{PR}[\mathbf{z}, k, t]$ leaks no partial information under $\mathrm{DPR}[n, k-1, t]$. ■

5.3 A Secure Stateful-Cipher

A cipher involves two parties, who share some common random input (the key). The goal of a cipher is the secure transmission of a sequence of messages. Suppose that I denotes the shared randomness between the sender and the receiver. A (stateful) cipher is defined by two probabilistic functions $f_I : \mathcal{K} \times \mathbb{P} \to \mathcal{K} \times \mathbb{C}$ and $g_I : \mathcal{K} \times \mathbb{C} \to \mathcal{K} \times \mathbb{P}$. The spaces $\mathcal{K}, \mathbb{P}, \mathbb{C}$ denote the state-space, plaintext-space and ciphertext-space respectively. The functions f, g have the property that if $f_I(s, m) = (s', c)$ (encryption) it holds that $g_I(s, c) = (s', m)$ (decryption); note that s' (given by both f, g) is the state that succeds the state s.

Stream-ciphers use public state sequences of the form $\langle 0, 1, 2, 3, \ldots \rangle$. The reader is referred to [Lub96] for more details on stream ciphers and how they can be built based on pseudorandom number generators. Block-ciphers encrypt messages of size equal to some fixed security parameter which are called blocks. Such ciphers are typically at the same state throughout and this state is considered to be secret (it coincides with the secret shared random key). The reader is referred to [Gol01] for further details on block-ciphers and generic constructions.

If a cipher, which operates on blocks, employs a "secret state-sequence update" and uses the shared randomness (the key) only as the initial state of the state-sequence, it is called a *stateful* cipher, see figure 1; (note that in a stateful cipher we suppress the subscript I from the functions f, g).

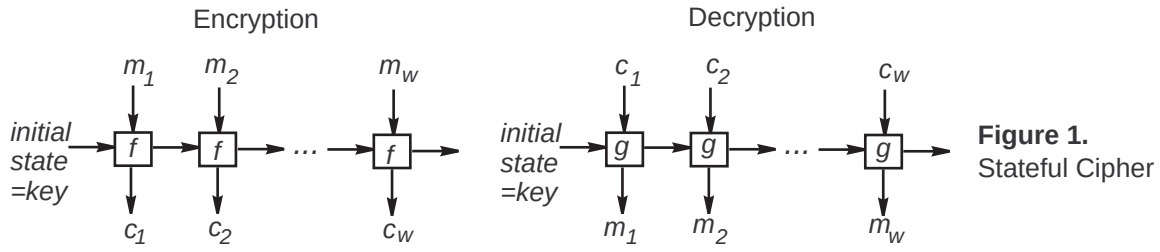

Figure 1: A Stateful Cipher

In the remaining of this section we introduce a stateful cipher that is based on PR and possesses a set of interesting properties: semantic security whose proof is derived from hardness of partial information extraction as established in section 4, random self-reducible properties that are based on the algebraic structure of the underlying problem and finally forward secrecy and error-correcting decryption which are easily demonstratable directly based on the construction.

5.3.1 Description of the PR-Cipher

The state-space $\mathcal{K}$ is defined to be the set of n-bitstrings with Hamming weight t, i.e., $\mathcal{K} = SS_{n,t}$. For some $s \in \mathcal{K}$ we define I_s to be the corresponding subset of $\{1, \ldots, n\}$, and v_s be the corresponding integer that has s as its binary representation. We denote by $V_{\mathcal{K}}$ the set of all numbers that their binary representation belongs in $\mathcal{K}$. Let $\mathbb{P} := \mathbb{F}^{\frac{k-1}{2}}$ and $\mathbb{C} := \mathbb{F}^n$. The shared randomness between the two parties is a random $s_0 \in \mathcal{K}$, that is the initial state of the

cipher. The encryption function of the cipher is defined as follows

$$f(s, \mathbf{m}) := F_{n,k,t}^{I_s}(\langle s', (\mathbf{m})_1, \ldots, (\mathbf{m})_{\frac{k-1}{2}} \rangle)$$

where $F_{n,k,t}^{I_s}$ is defined as the sampler S of definition 2.6 so that the subset I is set to I_s, the polynomial p is selected at random conditioned on $p(0) = s'$ and $p(i) = (\mathbf{m})_i$ for $i = 1, \ldots, \frac{k-1}{2}$. Note that s' is a random element of $V_{\mathcal{K}}$. The decryption function g is defined as follows: given $\langle s, C \rangle \in \mathcal{K} \times \mathbb{C}$, the polynomial p that corresponds to the pairs of C whose index is in I_s is interpolated. The decrypted message is set to be $\langle p(1), \ldots, p(\frac{k-1}{2}) \rangle$ and the next state is set to the binary representation of $p(0)$.

5.3.2 Semantic-Security

A semantic-security adversary $\mathcal{A}$ for a stateful cipher is a PPT that takes the following steps: (i) queries a polynomial number of times the encryption-mechanism (ii) generates two messages M_1, M_2 and obtain the ciphertext that corresponds to the encryption of M_b where b is selected at random from $\{1, 2\}$, (iii) queries the encryption-mechanism a polynomial number of times. Finally the adversary predicts the value of b. This is illustrated in figure 2. A cipher is said to be semantically secure if any semantic-security adversary predicts b with negligible advantage. For more details regarding semantically secure symmetric encryption, see [Lub96, KY00].

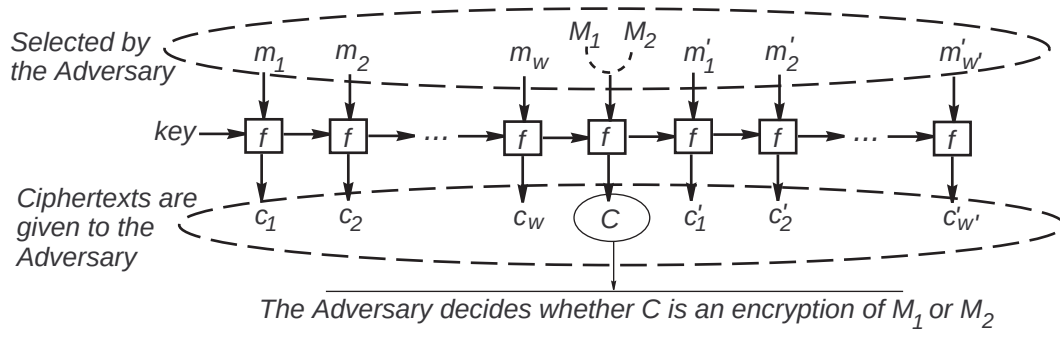

Figure 2. Semantic Security Adversary

Figure 2: Semantic security for stateful ciphers.

More formally semantic security in the context of stateful ciphers is defined as follows:

Definition 5.6 *Let $\mathcal{O}^b$, with $b \in \{1, 2\}$ be an encryption oracle for a stateful cipher initialized to a random initial state that accepts two kinds of input: (i) a plaintext, where $\mathcal{O}^b$ returns its encryption under the current state, (ii) a pair of plaintexts M_1, M_2, where $\mathcal{O}^b$ returns the encryption of M_b (such input is allowed only once). A semantic security adversary is a* PPT *$\mathcal{A}$ that given oracle access to $\mathcal{O}^b$ and attempts to predicts b (with probability better than $1/2$); the advantage of $\mathcal{A}$ is defined as follows:*

$$\mathsf{Adv}_{\mathsf{sem}}^{\mathcal{A}} = \mid \Pr[\mathcal{A}^{\mathcal{O}^b}(1^n) = b : b \leftarrow \{1, 2\}] - 1/2 \mid$$

where the probability is taken over all internal coin-tosses of $\mathcal{O}^b$ and $\mathcal{A}$ and all possible initial states for the cipher. If, for a certain cipher, there do not exist semantic security adversaries with non-nengligible advantage then we say that the cipher is semantically secure.

We remark that the two kinds of input to the encryption oracle define three stages of adversarial action, namely (i) querying the encryption oracle a number of times, (ii) submitting the "challenge" (the pair of plaintexts of which the adversary receives the encryption of one of the two at random), and (iii) querying the encryption oracle a number of times before guessing which of the two plaintexts of the challenge was encrypted. We proceed to show that the PR-Cipher is semantically secure under the Decisional PR-Assumption, specifically:

Theorem 5.7 *The PR-Cipher is semantically secure under the* DPR*: in particular for any $\mathcal{A}$ it holds that $\mathsf{Adv}^{\mathcal{A}}_{\mathsf{sem}} \leq (\mathsf{q}+1)\mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$ where q is the total number of queries posed by $\mathcal{A}$ to the encryption oracle ($\mathsf{q} = \mathrm{w} + \mathrm{w}'$ from figure 2).*

Proof. We start with a definition: we denote by $\mathcal{L}^{(u)}_{n,k,t}[\mathbf{m}_1, \ldots, \mathbf{m}_u]$ the vector output of an encryption oracle of the PR-cipher when accessed by a chosen plaintext adversary u times on messages $\mathbf{m}_1, \ldots, \mathbf{m}_u$. In other words it is the space of sequences of $\mathcal{I}_{\mathbf{z},k,t}$ instances $\mathbf{y}_1, \ldots, \mathbf{y}_u$ so that $\mathbf{m}_j := \langle p_{\mathbf{y}_j}(1), \ldots, p_{\mathbf{y}_j}(\frac{k-1}{2})\rangle$ and so that the binary representation of $p_{\mathbf{y}_j}(0)$ corresponds to the characteristic string of index set I of the tuple $\mathbf{y}_{j+1}$, for $j = 1, \ldots, u-1$. For two families of sets A_n and B_n we write $A \approx B$ if they are polynomial-time indistinguishable (see definition 3.1).

Claim 1. For any $u \geq 1$, $L^{(u)}_{n,k,t}[\mathbf{m}_1, \ldots, \mathbf{m}_u] \approx \mathbb{F}^n \times L^{(u-1)}_{n,k,t}[\mathbf{m}_2, \ldots, \mathbf{m}_u]$ under the assumption that $\mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$ is negligible.

Proof. Suppose the two families are distinguishable be some adversary $\mathcal{A}$ with non-negligible advantage. We will show how to use the adversary to violate the DPR with parameters $[n, (k-1)/2, t]$.

Adaptive Encryption Oracle. This is an oracle that has as input a tuple $\mathbf{y} \in \mathbb{F}^n$ such that $\mathbf{y} = \langle y_1, \ldots, y_n\rangle$ and $z_i \notin \{0, \ldots, k-1\}$ and receives u queries equal to the sequence of messages $\mathbf{m}_1, \ldots, \mathbf{m}_u$. Let $p'(x)$ be a polynomial so that (i) $p'(0)$ is a random element with $p'(0) \leq 2^n$ and the Hamming weight of $p'(0)$ is t, and (ii) $p'(i) = (\mathbf{m}_1)_i$ for $i = 1, \ldots, \frac{k-1}{2}$. Consider the transformation of $\mathbf{y}$ denoted by $\mathbf{y}_{m_1}$ and defined by $(\mathbf{y}_{m_1})_i = p'(z_i) + (\mathbf{y})_i \cdot \prod_{\ell=0}^{(k-1)/2}(z_i - \ell)$. Define I_2 to be the subset of $\{1, \ldots, n\}$ so that its characteristic string is identical to the binary representation of $p'(0)$. Next we sample $\mathbf{y}_{m_2}$ so that (i) $\langle p_{\mathbf{y}_{m_2}}(1), \ldots, p_{\mathbf{y}_{m_2}}(\frac{k-1}{2})\rangle = \mathbf{m}_2$, and (ii) the characteristic string of the index-set I for the instance $\mathbf{y}_{m_2}$ is identical to the binary representation of $p_{\mathbf{y}_{m_2}}(0)$. Continuing in a similar manner we construct the instances $\mathbf{y}_{m_1}, \ldots, \mathbf{y}_{m_u}$. It is clear that this series of samples is uniformly distributed over $L^{(u)}_{n,k,t}[\mathbf{m}_1, \ldots, \mathbf{m}_u]$ if the given $\mathbf{y}$ is drawn from $\mathsf{S}^{\mathbf{z}}_{n,(k-1)/2,t}$ whereas if $\mathbf{y}$ is drawn from $\mathbb{F}^n$ the output of the adaptive encryption oracle will be distributed uniformly over $\mathbb{F}^n \times L^{(u-1)}_{n,k,t}[\mathbf{m}_2, \ldots, \mathbf{m}_u]$. It follows that any distinguisher between the two distributions yields a distinguisher between random tuples of $\mathbb{F}^n$ and random instances of $\mathcal{I}_{\mathbf{z},(k-1)/2,t}$. This completes the proof of claim 1.

Claim 2. $L^{(u)}_{n,k,t}[\mathbf{m}_1, \ldots, \mathbf{m}_u] \approx (\mathbb{F}^n)^u$ under the assumption $u \cdot \mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$ is negligible.

Proof. Suppose that there is a distinguisher $\mathcal{A}$ between the two distributions. Then by the triangular inequality $\mathcal{A}$ can be used to distinguish two "neighboring hybrid distributions" that

are defined as follows: $(\mathbb{F}^n)^v \times L_{n,k,t}^{(u-v)}[\mathbf{m}_{u-v}, \ldots, \mathbf{m}_u]$, $(\mathbb{F}^n)^{v+1} \times L_{n,k,t}^{(u-v-1)}[\mathbf{m}_{u-v-1}, \ldots, \mathbf{m}_u]$ for some $v \in \{0, \ldots, u-1\}$. Using claim 1 we obtain the statement of the lemma.
(*Proof* of theorem 5.7) Suppose now that $\mathcal{A}$ is a semantic security adversary for the PR-cipher. Consider a predicate $\mathcal{B}$ that simulates $\mathcal{A}$ returning as ciphertexts random elements from $\mathbb{F}^n$ (including the ciphertext for the challenge step for which $\mathcal{B}$ flips the bit b). Finally $\mathcal{B}$ returns 1 if $\mathcal{A}$ guesses successfully the bit b. Using the result of claim 2 above it follows easily that the success probability of the adversary $\mathcal{A}$ will be bounded by $(\mathsf{q}+1)\mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$. ■

5.3.3 Random Self Reducibility Properties

In this section we study some basic random self-reducibility properties of the PR-cipher that are based on its underlying algebraic properties. We start with a definition: a chosen-plaintext-driven (cpd) adversary for a stateful cipher is a pair $\langle \mathcal{A}, \phi \rangle$ defined as follows:

Definition 5.8 *Let $\mathcal{E}_\kappa(m_\mathsf{c})$ be an encryption oracle for a stateful cipher that accepts two types of queries (posed in arbitrary order): (1) encryption queries $\langle \mathtt{encrypt}, m \rangle$ for which it returns the encryption of m under the current state, (2) a single challenge query $\langle \mathtt{challenge} \rangle$ for which it returns the encryption of m_c. A chosen-plaintext-driven (cpd) adversary with probability of success α is a pair $\langle \mathcal{A}, \phi \rangle$ where $\mathcal{A}$ is a* PPT *and ϕ is a polynomial-time computable function so that the following hold: $\mathsf{Pr}[\mathcal{A}^{\mathcal{E}_\kappa(m_\mathsf{c})}(1^n) = \phi(m_\mathsf{c}) : \kappa \leftarrow \mathcal{K}, m_\mathsf{c} \leftarrow \mathbb{P}] \geq \alpha$.*

Next we considerthe PR-cipher with parameters n, k, t. Below we define the following specialized versions of the above definition :

Definition 5.9 *For the PR-cipher with parameters n, k, t, we have*

1. Specialized Key Space dversary*: a specialized-key-space adversary is a specialized cpd-adversary $\langle \mathcal{A}, \phi \rangle$ together with a key distribution $\mathcal{D}$ that with probability α satisfies: $\mathsf{Pr}[\mathcal{A}^{\mathcal{E}_\kappa(m_\mathsf{c})}(1^n) = \phi(m_\mathsf{c}) : \kappa \leftarrow \mathcal{D}, m_\mathsf{c} \leftarrow \mathbb{P}] \geq \alpha$.*

2. Specialized Plaintext Space Adversary*: a specialized-plaintext-space adversary is a specialized cpd-adversary $\langle \mathcal{A}, \phi \rangle$ together with a plaintext distribution $\mathcal{D}$ that with probability α satisfies: $\mathsf{Pr}[\mathcal{A}^{\mathcal{E}_\kappa(m_\mathsf{c})}(1^n) = \phi(m_\mathsf{c}) : \kappa \leftarrow \mathcal{K}, m_\mathsf{c} \leftarrow \mathcal{D}] \geq \alpha$.*

3. Partial-Domain Adversary*: a partial-domain cpd adversary $\langle \mathcal{A}, \phi \rangle$ with probability α sastisfies that $\phi \in \{\mathsf{Proj}_1, \ldots, \mathsf{Proj}_{\frac{k-1}{2}}\}$ where $\mathsf{Proj}_i : \mathbb{F}^{\frac{k-1}{2}} \rightarrow \mathbb{F}$ such that $\mathsf{Proj}_i(\langle m_1, \ldots, m_{\frac{k-1}{2}} \rangle) = m_i$.*

In this section, we will employ an extended formulation of the PR-cipher that chooses a different support vector $\mathbf{z}$ for each ciphertext that is selected at random from $\mathbb{F}^n$ under the constraint that all coordinates are distinct elements of $\mathbb{F}$ and none of them is included to the set $\{0, 1, \ldots, k\}$. We remark that all the results of the previous sections regarding the PR-cipher carry very easily to the above modification that has a randomized support vector as in all the arguments no particular property of the support vector was used (note that this would also require that the intractability assumption should be amended to quantify over a random choice of the support vector as well). We call the modified cipher, the PR-cipher with randomized support. The random self-reducibility properties of the PR-cipher with randomized support are established in the following theorem:

Theorem 5.10 *For the PR-cipher with randomized support it holds that:*

1. *A cpd-adversary $\langle \mathcal{A}, f\rangle$ with probability of success α can be transformed to a specialized-key-space adversary $\langle \mathcal{A}, f\rangle$ for any key distribution $\mathcal{D}$ and the same probability α.*
2. *A cpd-adversary $\langle \mathcal{A}, f\rangle$ with probability α can be transformed to a specialized-plaintext-space cp-adversary $\langle \mathcal{A}, f\rangle$ for any plaintext distribution $\mathcal{D}$ and the same probability α, provided that $\phi : \mathbb{F}^{\frac{k-1}{2}} \rightarrow G$ where $(G, *)$ is a group and ϕ is a group homomorphism from $(\mathbb{F}^{\frac{k-1}{2}}, +)$ to $(G, *)$.*
3. *A partial-domain cpd-adversary $\langle \mathcal{A}, \mathsf{Proj}_j\rangle$ with probability α can be transformed to a partial-domain cpd-adversary $\langle \mathcal{A}, \mathsf{Proj}_{j'}\rangle$ with probability $\alpha - \mathsf{negl}(\lambda)$ for any $j' \neq j$, provided that $n \cdot k \leq \epsilon \cdot |\mathbb{F}|$ where $\epsilon \in (0, 1)$.*

Proof. (1) The proof follows easily from the fact that in case we have a randomized support, any ciphertext $(\mathbf{z}, \mathbf{y})$ can be permuted according to a random permuation π, an operation that will effectively randomize the secret-key used (which corresponds to the error-locations) independently of the key distribution $\mathcal{D}$ where the secret-key is drawn from. As a result the definition of the specialized-key-space cpd-adversary $\mathcal{A}'$ is simply to select π at random and apply the permutation to randomly shuffle the elements of the two challenge $(\mathbf{z}, \mathbf{y})$; then $\mathcal{A}'$ simulates the cpd-adversary $\mathcal{A}$ and returns its output. The statement of the theorem follows easily.

(2) Similarly as in the case of (1) we observe that we can randomize the plaintext of the challenge ciphertext $(\mathbf{z}, \mathbf{y})$ by selecting a random polynomial $R(x)$ of degree less than k and computing $(\mathbf{z}, \mathbf{y} + \mathbf{R})$. As before the specialized-plaintext-space cpd-adversary can apply this transformation to the challenge ciphertext and then feed this challenge to the cpd-adversary $\mathcal{A}$. Given the output of $\mathcal{A}$ we observe that with probability α it equals $\phi(m_{\mathsf{c}} + R)$ where m is a $\frac{k-1}{2}$-vector over $\mathbb{F}$ and R is the vector $\langle R(1), \ldots, R(\frac{k-1}{2})\rangle$. Based on the homomorphic property we have that $\phi(m_{\mathsf{c}} + R) = \phi(m_{\mathsf{c}}) * f(R)$ and from this the specialized-plaintext-space cpd-adversary $\mathcal{A}'$ will be able to recover $\phi(m_{\mathsf{c}})$ by dividing by $f(R)$. Note that the randomized support is not essential here (and thus this result will carry to the basic PR-cipher as well).

(3) Suppose that $\mathcal{A}$ is a partial-domain cpd-adversary for Proj_j where $j \in \{1, \ldots, \frac{k-1}{2}\}$ and let $j' \in \{1, \ldots, \frac{k-1}{2}\}$ such that $j \neq j'$. We know that $\mathcal{A}$ returns $\mathsf{Proj}_j(m_{\mathsf{c}})$ and we want to transform $\mathcal{A}$ to an algorithm $\mathcal{A}'$ that returns $\mathsf{Proj}_{j'}(m_{\mathsf{c}})$. Consider a PR-instance $\mathbf{y}$ with support $\mathbf{z}$. Now suppose that we consider $\mathbf{y}$ to have support $a \cdot \mathbf{z} + b$ where $a, b \in \mathbb{F}$. It follows that if p is the polynomial solution of $\mathbf{y}$ based on support $\mathbf{z}$ then $p((x - b)/a)$ would be the polynomial solution of $\mathbf{y}$ based on support $a \cdot \mathbf{z} + b$. Given that $\mathcal{A}$ can recover $p(j)$ we simply need to select a, b such that $(j - b)/a = j'$ and simulate $\mathcal{A}$ on challenge $\mathbf{y}$ with support $a \cdot \mathbf{z} + b$. Suppose we select b at random from $\mathbb{F}$ and $a = (j' - b)/j$. Define the event BAD_i to be the event that $az_i + b \in \{0, 1, \ldots, k-1\}$ for the above selection of a, b; it is easy to see that $\mathsf{Pr}[\mathsf{BAD}_i] \leq \frac{k}{|\mathbb{F}|}$ and as a result the probability $\mathsf{Pr}[\cup_i \mathsf{BAD}_i] \leq n\frac{k}{|\mathbb{F}|}$. Assuming that $n(k + 1) \leq \epsilon \cdot |\mathbb{F}|$ where $\epsilon \in (0, 1)$ is a constant it follows that the choice a, b will be appropriate with error probability at most ϵ. We may repeat the sampling of b a number of q times independently to reduce the error to $\epsilon^q = \mathsf{negl}(\lambda)$ and as a result conclude that the success probability of $\mathcal{A}'$ will be $\alpha - \mathsf{negl}(\lambda)$. ■

We conclude with the following corollary that shows that any attack on the average-case of the plaintext space or the key-space is equivalent to an attack to any specific restriction of the key-space or plaintext-space which is large enough.

Corollary 5.11 *For the PR-cipher with parameters n, k, t based on λ the following statements are equivalent:*

1. *There exists a cpd-adversary with non-negligible probability of success.*
2. *There exists a specialized key-space adversary for some key space $\mathcal{K}' \subseteq \mathcal{K}$ with $\#\mathcal{K}'/\#\mathcal{K}$ non-negligible in λ where the distribution $\mathcal{D}$ is defined as the uniform distribution over $\mathcal{K}'$.*
3. *There exists a specialized plaintext-space adversary for some plaintext space $\mathbb{P}' \subseteq \mathbb{P}$ with $\#\mathbb{P}'/\#\mathbb{P}$ non-negligible in λ where the distribution $\mathcal{D}$ is defined as the uniform distribution over $\mathbb{P}'$.*

Proof. The implications (1) $\Rightarrow$ (2) and (1) $\Rightarrow$ (3) follow from theorem 5.10. The reverse implications are immediate due to the size of the underlying restricted spaces. ■

5.3.4 Forward Secrecy

A cipher is said to satisfy forward secrecy if in the case of a total security breach at some point of its operation (i.e. the internal state is revealed) the adversary is unable to extract any information about the previously communicated messages.

This is formalized by an adversary that mounts two chosen plaintext security attacks submitting adaptively messages to the encryption oracle. The encryption oracle flips a coin and answers by encrypting the plaintexts submitted by one of the two attacks (the same attack is answered throughout). At some point the internal state of the system is revealed to the adversary. Forward secrecy is violated if the adversary can tell with probability significantly better than one half to which chosen plaintext attack is the encryption oracle responding. If the adversary cannot predict this with probability significantly better than 1/2 the a cipher is said to satisfy forward secrecy. More formally,

Definition 5.12 *Let $\mathcal{O}^b_{\text{fs}}$, with $b \in \{1, 2\}$ be an encryption oracle for a stateful cipher initialized to a random initial state that accepts two kinds of input: (i) a pair of plaintexts $\mathbf{m}_1, \mathbf{m}_2$, where $\mathcal{O}^b_{\text{fs}}$ returns the encryption of $\mathbf{m}_b$ under the current state, (ii) a termination message, where $\mathcal{O}^b_{\text{fs}}$ returns the current internal state; no more queries are accepted by $\mathcal{O}^b_{\text{fs}}$ after the termination message is submitted. A forward secrecy adversary is a* PPT *$\mathcal{A}$ that given oracle access to $\mathcal{O}^b_{\text{fs}}$ it attempts to predicts b (with probability better than $1/2$); the advantage of $\mathcal{A}$ is defined as follows:*

$$\mathsf{Adv}^{\mathcal{A}}_{\text{fs}} = \mid \mathsf{Pr}[\mathcal{A}^{\mathcal{O}^b_{\text{fs}}}(1^n) = b : b \leftarrow \{1, 2\}] - \frac{1}{2} \mid$$

where the probability is taken over all internal coin-tosses of $\mathcal{O}^b_{\text{fs}}$ and $\mathcal{A}$ and all possible initial states for the cipher. If, for a certain cipher, there do not exist forward secrecy adversaries with non-negligible advantage then we say that the cipher satisfies forward secrecy.

Below we establish the forward secrecy of the PR cipher based on the DPR assumption.

Theorem 5.13 *The PR-Cipher satisfies forward secrecy under* DPR*: in particular, for any* $\mathcal{A}$ *it holds that* $\mathsf{Adv}^{\mathcal{A}}_{\mathsf{fs}} \leq \mathsf{q} \cdot \mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$ *where* q *is the total number of queries posed by* $\mathcal{A}$ *to the encryption oracle.*

Proof. We denote by $\mathcal{L}^{(u)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_1,\ldots,\mathbf{m}^0_u\\ \mathbf{m}^1_1\end{smallmatrix}\!\!\begin{smallmatrix}\\ \mathbf{m}^1_u\end{smallmatrix}]$ the output of an encryption oracle of the stateful cipher when accessed by the two chosen plaintext attacks that are part of the forward secrecy adversary. In other words it is the space of sequences of $\mathcal{I}_{\mathbf{z},k,t}$ instances $\mathbf{y}_1,\ldots,\mathbf{y}_u$ so that $\langle p_{\mathbf{y}_j}(1),\ldots,p_{\mathbf{y}_j}(\frac{k-1}{2})\rangle = \mathbf{m}^b_j$ for all $j = 1,\ldots,u$ where b is a random coin toss; the binary representation of $p_{\mathbf{y}_j}(0)$ corresponds to the characteristic string of the index set $\mathbf{y}_{j+1}$, for $j = 1,\ldots,u-1$.

Claim 3. For any $u \geq 1$, $\mathcal{L}^{(u)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_1,\ldots,\mathbf{m}^0_u\\ \mathbf{m}^1_1\end{smallmatrix}\!\!\begin{smallmatrix}\\ \mathbf{m}^1_u\end{smallmatrix}] \approx \mathbb{F}^n \times \mathcal{L}^{(u-1)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_2,\ldots,\mathbf{m}^0_u\\ \mathbf{m}^1_2\end{smallmatrix}\!\!\begin{smallmatrix}\\ \mathbf{m}^1_u\end{smallmatrix}]$ unless the DPR fails. In particular, the distance between the two distributions is at most $\mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$.

The arguments of the proof of claim 3 are very similar to those of the proof of claim 1 in the proof of theorem 5.7.

Claim 4. $\mathcal{L}^{(u)}_{n,k,t}[\begin{smallmatrix}\mathbf{m}^0_1,\ldots,\mathbf{m}^0_u\\ \mathbf{m}^1_1\end{smallmatrix}\!\!\begin{smallmatrix}\\ \mathbf{m}^1_u\end{smallmatrix}] \approx (\mathbb{F}^n)^u$ unless the DPR fails. In particular, the distance between the two distributions is at most $u \cdot \mathsf{Adv}^{\mathsf{psr}}_{\mathbf{z},(k-1)/2,t}$.

The arguments of the proof of claim 4 are very similar to those of the proof of claim 2 in the proof of theorem 5.7.

Based on the above, it follows that we may simulate the ciphertexts provided to the adversary and the statement of the theorem follows easily as in the case of theorem 5.7: for any choice of b the output of the encryption oracle is independent of the sequence of simulated ciphertexts submitted for encryption by the oracle. ■

5.3.5 Error-Correcting Decryption

A cryptosystem is said to allow error-correcting decryption if the decryption procedure is able to correct errors that are introduced during the transmission (possibly by an adversary). This combines the decryption operation with the error-correction operation.

A cryptosystem that transmits plaintext blocks of size d is called d'-error-correcting if up to d' errors can be corrected The PR-cipher (which transmits plaintexts blocks of size $\frac{k-1}{2}$ over the underling finite field $\mathbb{F}$) is error-correcting since the interpolation step during decryption can be substituted by the [BW86] polynomial-reconstruction algorithm that can withstand up to $\frac{t-k}{2}$ errors (in the worst-case). Extended error-correction capability can be achieved if the Guruswami [GS98] list-decoder is applied instead of the [BW86] decoding method (but in this case decryption may not be unique).

6 Conclusion

In this work we layed out a framework for the employment of the Polynomial Reconstruction problem in cryptography. We put forth a natural decisional intractability assumption that appears to be intimately related to the decoding problem: distinguishing a randomly chosen error location from a randomly chosen correct location of the codeword in a uniform noise setting. We showed that this assumption is sufficiently powerful to imply the pseudorandomness of PR instances, i.e., the indistinguishability of codewords that are hard to decode from purely

random vectors. Furthermore we established the fact that under our decisional assumption PR-instances leak no partial information for a number of points of their polynomial solution, i.e., a PR-instance semantically hides a number of solution points. Based on these results we showed three cryptographic applications: (i) pseudorandom number generation, (ii) oblivious polynomial evaluation, (iii) a semantically secure stateful-cipher.

References

[BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. SIAM Journal on Computing, 26(5):1510—1523, October 1997.

[Ber68] Elwyn R. Berlekamp, *Algebraic Coding Theory.* McGraw-Hill, 1968.

[BW86] Elwyn R. Berlekamp and L. Welch, *Error Correction of Algebraic Block Codes.* U.S. Patent, Number 4,633,470 1986.

[BKY03] Daniel Bleichenbacher, Aggelos Kiayias and Moti Yung, Decoding of Interleaved Reed Solomon Codes over Noisy Data, ICALP 2003, pp. 97-108.

[BG84] Manuel Blum and Shafi Goldwasser, *An Efficient Probabilistic Public-Key Encryption Scheme Which Hides All Partial Information*, Crypto 1984.

[CS03] Don Coppersmith and Madhu Sudan, Reconstructing curves in three (and higher) dimensional space from noisy data, STOC 2003, pp. 136-142.

[Gol90] Oded Goldreich, *A note on computational indistinguishability*, Information Processing Letters, vol. 34, no. 6, pp. 277–281, 1990.

[Gol93] Oded Goldreich, *A Uniform Complexity Treatment of Encryption and Zero-Knowledge*, Journal of Cryptology, Vol. 6, pp.21-53, 1993.

[Gol01] Oded Goldreich, *Foundations of Cryptography*, Cambridge University Press. 2001.

[GSR95] Oded Goldreich, Madhu Sudan and Ronitt Rubinfeld, *Learning Polynomials with Queries: The Highly Noisy Case.* In the Proceedings of the 36th Annual Symposium on Foundations of Computer Science, 1995.

[GM84] Shafi Goldwasser and Silvio Micali, *Probabilistic encryption*, Journal of Computer and System Sciences, vol. 28(2), pp. 270-299, April 1984.

[GS98] Venkatesan Guruswami and Madhu Sudan, *Improved Decoding of Reed-Solomon and Algebraic-Geometric Codes.* In the Proceedings of the 39th Annual Symposium on Foundations of Computer Science, 1998.

[HILL99] Johan Hastad, Russel Impagliazzo, Leonid Levin and Michael Luby, *Construction of a pseudo-random generator from any one-way function*, SIAM J. Comput. 28(4):1364-1396, 1999.

[KY02] Aggelos Kiayias and Moti Yung, *Cryptographic Hardness based on the Decoding of Reed Solomon Codes*, In the Proceedings of the 29th International Colloquium in Algorithms, Languages and Programming, 2002.

[KY04] Aggelos Kiayias and Moti Yung,*Directions in Polynomial Reconstruction Based Cryptography*, IEICE Transactions, Vol. E87-A, No. 5, May 5, 2004. pp. 978–985.

[KY00] Jonathan Katz and Moti Yung, *Complete Characterization of Security Notions for Probabilistic Private-key Encryption*, In the Proceedings of the 32th ACM Symposium on the Theory of Computing, 2000.

[Lub96] Michael Luby, *Pseudorandomness and Cryptographic Applications*, Princeton University Press, 1996.

[MS77] F. J. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes.* North Holland, Amsterdam, 1977.

[Na91] Moni Naor, *Bit Commitment Using Pseudorandomness*, Journal of Cryptology 4(2): pp. 151-158, 1991.

[NP99] Moni Naor and Benny Pinkas, *Oblivious Transfer and Polynomial Evaluation.* In the Proceedings of the 31th ACM Symposium on the Theory of Computing, 1999.

[NP06] Moni Naor and Benny Pinkas, *Oblivious Polynomial Evaluation.* SIAM J. Comput. 35(5), pp. 1254-1281, 2006.

[NR98] Moni Naor and Omer Reingold, *Number-theoretic Constructions of Efficient Pseudorandom Functions*, FOCS 1997.

[PV05] Farzad Parvaresh and Alexander Vardy, Correcting Errors Beyond the Guruswami-Sudan Radius, Foundations of Computer Science, 2005.

[Ped92] Torben P. Pedersen, *Non-Interactive and information-theoretic secure verifiable secret-sharing*, Crypto 1991.

[RS60] Irvin Reed and Gustave Solomon, "Polynomial codes over certain finite fields," SIAM Journal of Applied Mathematics, vol. 8, no. 2, pp. 300–304, 1960.

[Sud97] Madhu Sudan, *Decoding of Reed Solomon Codes beyond the Error-Correction Bound.* Journal of Complexity 13(1), pp. 180–193, 1997.

[Yao82] Andrew C. Yao, *Theory and applications of trapdoor functions*, In the Proceedings of IEEE Symposium on Foundations of Computer Science, pp. 80-91, 1982.